Structural Insights into Virulence Regulation in *Yersiniae* and Polyketide Synthase Systems

Von der Fakultät für Lebenswissenschaften

der Technischen Universität Carolo-Wilhelmina

zu Braunschweig

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

genemigte

D i s s e r t a t i o n

von Nick Quade
aus Hannover

1. Referent:	Honorarprofessor Dr. Dirk Heinz
2. Referent:	Professor Dr. Michael Steinert
eingereicht am:	04.05.2011
mündliche Prüfung (Disputation) am:	18.08.2011

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Publikation

Quade, N. Dieckmann, M., Haffke, M., Heroven, A.K. Dersch, P. & Heinz, D.W.: Structure of the effector-binding domain of the LysR-type transcription factor RovM from *Yersinia pseudotuberculosis*. Acta Crystallographica Section D 67, 81-90 (2011).

Tagungsbeiträge

Quade, N., Grujic, B., Herbst, K.. Dersch, P. & Heinz, D.W.: Structure of the transcription factor RovA from *Yersinia pseudotuberculosis*. (Vortrag) 3rd International PhD symposium of the Helmholtz International Research school, HZI, Braunschweig, Dezember 2009

Quade, N., Grujic, B., Herbst, K.. Dersch, P. & Heinz, D.W.: Structure of the transcription factor RovA from *Yersinia pseudotuberculosis*. (Poster) Monod symposium: Gene expression and signalling in bacteria, Institut Pasteur, Paris, May 2010

Quade, N., Grujic, B., Herbst, K.. Dersch, P. & Heinz, D.W.: Structure of the transcription factor RovA from *Yersinia pseudotuberculosis*. (Vortrag) 13th Heart of Europe Bio-Crystallography Meeting, Schöneck, Vogtland, Germany, September 2010

Quade, N., Grujic, B., Herbst, K.. Dersch, P. & Heinz, D.W.: Structure of the transcription factor RovA from *Yersinia pseudotuberculosis*. (Poster) - 1st North-Regio-Day on Infection, HZI, Braunschweig, Oktober 2010

Quade, N., Grujic, B., Herbst, K.. Dersch, P. & Heinz, D.W.: Structure of the transcription factor RovA from *Yersinia pseudotuberculosis*. (Poster) - 10th International Symposium on Yersinia, Recife, Brazil, Oktober 2010, 1st poster prize

Quade, N., Rashid, S. Huo, L. Müller, R. & Heinz, D.W.: New ways in carbon fixation - Structural studies of the Octenoyl-CoA reductase/carboxylase CinF. (Vortrag) 4th International PhD symposium of the Helmholtz International Research school, HZI, Braunschweig, Dezember 2010

Contents

Abbreviations

Å	Ångström
ACP	Acyl carrier protein
AT	Acyltransferase
ATP	Adenosine triphosphate
CCP4	Collaborative Computational Project No. 4
CCR	Crotonyl-CoA carboxylase/reductase
CD	Circular dichroism
CoA	Coenzyme A
Csr	Carbon storage regulator
DBD	DNA binding domain
DEBS	6-deoxyerythronolide B synthase
DH	Dehydrogenase
EBD	Effector binding domain
EM	Extension module
ER	Enoylreductase
ERSF	European Synchrotron Radiation Facility
ESI	Electrospray Ionisation
FAS	Fatty acid synthase
H-NS	Histone-like nucleoid structuring protein
IPTG	Isopropyl-β-D-thiogalactopyranosid
KS	Ketosynthase
KR	Ketoreductase
LD_{50}	Lethal dose for 50% deaths
LM	Loading module
LTTR	LysR-type transcription regulator
LPS	Lipopolysaccharide
MAD	Multiple wavelength anomalous dispersion
MALDI-TOF	Matrix Assisted Laser Desorption Ionisation – Time Of Flight
MIR	Multiple isomorphous replacement
MR	Molecular replacement
MRSA	Methicillin-resistant *Staphylococcus aureus*

MS	Mass spectrometry
MWCO	Molecular weight cutoff
NAD(H)	Nicotinamide adenine dinucleotide
NADP(H)	Nicotinamide adenine dinucleotide phosphate
Ni-NTA	Nickel-nitrilotriacetic acid
NRPS	Non-ribosomal peptide synthetase
OD_{600}	Optical density at wavelengh of 600nm
PCR	Polymerase chain reaction
PDB	Protein database
PEG	Polyethylene glycol
PIK	Pikromycin
PKS	Polyketide synthase
PPant	Phosphopantetheine
rmsd	Root main square deviation
RovA	Regulator of virulence activator
RovM	Regulator of virulence modulator
rpm	Rounds per minute
SAD	Single wavelength anomalous dispersion
SD200	Superdex 200 (gel filtration column)
SDS	Sodium dodecyl sulfate
SDS PAGE	SDS polyacrylamide gel electrophoresis
SeMet	Selenium methionine
SIR	Single isomorphous replacement
SpirTE	Spirangien thioesterase
TE	Thioesterase
TLS	Translation liberation screw
TMC	Tautomycetin
V_M	Matthews coefficient
wHTH	winged helix-turn-helix
YadA	*Yersinia* adhesin A
Ysc	*Yersinia* secretion
Yop	*Yersinia* outer protein

Zusammenfassung

Teil A: Virulenzregulation in pathogenen *Yersinien*

Enteropathogene Yersinien verwenden ein Adhesin namens Invasin, um die Darmepithel zu überwinden und in den Körper des Wirts zu gelangen. Dieses Protein wird von dem Transkriptionsfaktor RovA reguliert. RovA fungiert als zentrale Schaltstelle in einem komplizierten Regulationsnetzwerk, das mehrere Proteine und RNAs umfasst. Besonders interessant ist seine Fähigkeit als Proteinthermometer zu fungieren. Das bedeutet, dass RovA bei erhöhter Temperatur eine Konformationsumwandlung durchläuft, welche zu niedrigerer DNA-Bindungsaffinität und verstärktem Abbau durch Proteasen führt.

In dieser Arbeit wurde RovA strukturell und biochemisch untersucht, um ein besseres Verständnis dieses komplexen Regulationsnetzwerkes zu entwickeln. Die Struktur von RovA im Komplex mit DNA wurde bei 1.9 Å Auflösung gelöst. Diese zeigt nur wenige basenspezifische Interaktionen. Dies erklärt die geringe Selektivität von RovA für die DNA und somit wie RovA an unterschiedliche DNA-Sequenzen im Genom binden kann. Weiterhin wurde entdeckt, dass mehrere kleine Moleküle wie z.B. Salicylsäure mit millimolarer Affinität an RovA binden können, was zu reduzierter DNA-Bindeaffinität und temperaturabhängiger Konformationsumwandlung führt. Um diese Effekte auf molekularer Ebene verstehen zu können, wurde auch die Struktur von RovA im Komplex mit Salicylsäure gelöst, in der zwei Bindestellen für Salicylsäure pro Monomer RovA identifiziert werden konnten. Die Bindung von Salicylsäure führt zu einer Konformationsänderung gegenüber der DNA-gebundenen Form. Dies erklärt, warum die DNA-Affinität von RovA mit Salicylsäure niedriger ist.
Darüberhinaus wurde in dieser Arbeit noch RovM, ein weiteres Protein aus diesem Regulationsnetzwerk biophysikalisch untersucht. RovM ist ein LysR-ähnlicher Transkriptionsfaktor, der in der RovA Promoterregion bindet und dadurch die Produktion von RovA verhindert. Es konnte gezeigt werden, dass Volllänge RovM ein Tetramer bildet, während die Effektorbindedomäne von RovM (RovM-EBD) in Lösung als Dimer vorliegt. Die Struktur von RovM-EBD wurde gelöst, welche aus zwei Domänen besteht. Zwischen diesen befindet sich eine Öffnung, in der möglicherweise ein Induktormolekül binden kann.

Teil B: Polyketidsynthasen

Der zweite Teil dieser Arbeit behandelt die Strukturaufklärung von Proteinen, die an der Synthese von Polyketiden beteiligt sind. Polyketide sind Naturstoffe, die viele Anwendungen in der Medizin finden, z.B. als Antibiotika. Sie werden von riesigen Multienzymkomplexen namens Polyketidsynthasen hergestellt, die die Polyketide wie am Fließband schrittweise aus kleinen Bausteinen, sogenannte Erweiterungseinheiten, zusammensetzen.

Solche Erweiterungseinheiten werden oft von speziellen Enzymen, sogenannten Crotonyl-CoA Carboxylasen/Reduktasen (CCR), produziert. Diese katalysieren die reduktive Carboxylierung von Crotonyl-CoA zu der typischen Erweiterungseinheit Ethylmalonyl-CoA. Allerdings gibt es auch Proteine, die andere, ungewöhnliche Erweiterungseinheiten herstellen, wie z.B. Hexylmalonyl-CoA. In dieser Arbeit wurde eine solche Karboxylase/Reduktase (CinF), die die Konvertierung von Oktenoyl-CoA zu Hexylmalonyl-CoA katalysiert, strukturell und biochemisch untersucht. Die Struktur von CinF wurde alleine und im Komplex mit den Liganden NADP und Oktenoyl-CoA gelöst. So war es möglich, einen Reaktionsmechanismus für diese ungewöhnliche Karboxylierungsreaktion zu postulieren und Aminosäuren zu bestimmen, die für die Auswahl der Substrate zuständig sind. Dadurch könnte es möglich werden, diese Proteine umzuprogrammieren, sodass sie andere Erweiterungseinheiten bereitstellen. Diese könnten dann verwendet werden, um neuartige Medikamente mit verbesserten Eigenschaften herzustellen.

Weiterhin wurde die Thioesterase-Domäne aus dem Spirangien PKS Biosyntheseweg (SpirTE) kristallisiert. Während der Produktion bleiben die Polyketide die ganze Zeit kovalent über eine Thioesterbindung an die PKS-Proteine gebunden. Die TE-Domäne katalysiert die Abspaltung des fertigen Produkts vom Protein im letzten Schritt, wobei es häufig auch zu einer Zyklisierung des Produkts kommt. SpirTE hingegen setzt ein lineares Produkt frei. Das Ziel dieses Projekts ist es, auf struktureller Ebene zu erklären, warum manche TE-Domänen eine Zyklisierung bewirken und andere nicht. Die Struktur von SpirTE zeigt einen Substrattunnel, der sich durch das ganze Protein zieht und in der Mitte das aktive Zentrum beherbergt. Verglichen mit anderen TE-Domänen ist dieser Tunnel deutlich schmaler, was verhindert, dass das Produkt zyklisiert.

Summary

Part A: Virulence regulation in pathogenic *Yersiniae*

Enteropathogenic Yersiniae use an adhesin called invasin for initial penetration of the intestinal barrier to reach the interior of the host. This protein, among many others, is regulated by the global transcription factor RovA. It is the major player in an intricate regulatory network comprising several proteins and RNAs. An interesting observation was that RovA could act as a protein thermometer. This means that it undergoes a conformational change in response to an increase in temperature which leads to reduced DNA binding affinity and increased susceptibility to proteolytic degradation.

In this work, structural and biochemical methods were used to gain a deeper insight into the complex regulation by RovA. The structure of RovA in complex with DNA was solved at 1.9 Å resolution showing how RovA interacts with the DNA. The few specific interactions with the bases of the DNA explain how RovA can bind to several different binding sites in the bacterial genome. Several small molecules such as salicylate were found to bind to RovA with millimolar affinity and reduce its DNA binding affinity as well as its temperature dependent conformational change. In order to understand these effects on a molecular level, the structure of RovA in complex with salicylate was solved at 2.4 Å resolution. This showed two salicylate molecules bound per RovA monomer. Salicylate binding leads to a small conformational change compared to the DNA bound structure which might account for the reduced DNA binding affinity in the presence of salicylate.

Additionally, another protein from this regulatory network, RovM, was analysed biophysically in this work. RovM is a LysR-type transcription factor that binds to the RovA promoter and thereby inhibits RovA production. It was possible to show that full length RovM is a tetramer in solution, whereas its effector binding domain (RovM-EBD) exists as a dimer. Furthermore, the structure of RovM-EBD was determined at 2.4 Å resolution comprising two domains. A cavity was detected between the two domains that could bind a potential inducer molecule.

Part B: Polyketide synthases

The second topic of this thesis deals with the structural investigation of proteins involved in polyketide synthesis. Polyketides are natural products, often from soil bacteria, that have many applications in medicine for example as antibiotics. They are produced by huge multi-enzyme complexes called polyketide snythases (PKS). These assemble polyketides in a step by step manner from small building blocks, similar to an assembly line.

The building blocks, called extender units, are often generated by dedicated enzymes, called crotonyl-CoA carboxylase/reductases (CCR). These catalyse the reductive carboxylation of crotonyl-CoA to yield the typical extender unit ethyl-malonyl-CoA. However, recently carboxylase/reductases have been described that produce other, unusual building blocks such as hexyl-malonyl-CoA or chloro-ethyl-malonyl-CoA. In this work, structural and biophysical analysis of such an unusual carboxylase/reductase enzyme (CinF) that catalyses the conversion of octenoyl-CoA to hexyl-malonyl-CoA has been undertaken. The structure of CinF was solved alone and in complex with its ligands NADP and octenoyl-CoA. This represents the first structure of such an enzyme in complex with its ligands. It was possible to propose a reaction mechanism for this unusual carboxylation reaction and to determine amino acids that are involved in the selection of the substrate. Therefore it might become possible to reprogram carboxylase/reductases to process novel extender units which could be incorporated into new drugs with enhanced properties.

Additionally, the thioesterase domain from the spirangien PKS biosynthesis cluster (SpirTE) was crystallized. During polyketide synthesis the polyketide stays covalently bound to the PKS proteins at all times via a thioester. TE domains catalyse the last step in polyketide synthesis where they cleave this thioester link to release the final product and often also catalyse the cyclisation of the product. SpirTE, on the other hand, releases a linear product. The aim of this project was to understand on a structural level why some TE domains catalyse cyclisation of the product and some do not. The structure of SpirTE showed a substrate tunnel spanning the whole of the protein which harbours the active site in the middle. Compared to structures of TE domains which catalyse cyclisation SpirTE has a much more constricted substrate tunnel which could prevent the looping of the substrate and thereby its cyclisation.

1 Introduction

1A Virulence regulation in pathogenic *Yersiniae*

1A.1 The plague

The genus *Yersinia* is very infamous due to its role in the plague pandemics which killed perhaps up to 100 million people worldwide in the 14th century. Originating in central Asia, the pandemic, then called Black Death, spread throughout Asia, Europe and Africa, reducing the population of Europe by 1/3 with substantial impact on European politics, culture and religion (Stenseth *et al.* 2008). However, there had been severe outbreaks of the plague before that, e.g. the so called Justinian plague in the 6th century, which killed about half of the population in Europe and is thought to have facilitated the Arab conquest of North Africa and Spain. Some sources even describe outbreaks of the plague in Greece possibly caused by *Yersinia* infections that date back to 430 BC (Perry & Fetherston 1997).

Though today the hygienic conditions in most parts of the world have improved substantially and the disease is treatable with antibiotics, the plague is still endemic in some parts of the world, mostly Africa. There it continues to be a threat with several thousand reported cases every year. Especially worrying is the emergence of multi-drug resistance (Galimand *et al.* 1997). The plague has also been tested as a biological weapon by the Japanese Army in World War II (Harris 1994). Aerosol dissemination of the plague in densely populated areas for example by a terrorist attack would lead to over 100,000 infections and several 10,000 fatalities (Inglesby *et al.* 2000). Thus, even today the plague remains a considerable concern.

1A.2 *Yersinia*

The plague is caused by a member of the bacterial genus *Yersinia*, called *Yersinia pestis*. The genus *Yersinia* is part of the family of Enterobacteriacea and was named after the Swiss bacteriologist Alexandre Jean Emile Yersin (1863-1943) who

discovered *Yersinia pestis* as the causative agent of the bubonic plague (Perry & Fetherston 1997). *Yersinia* bacteria are facultative anaerobic, Gram-negative rod shaped bacteria with a length of a few micrometers (Fig. 1A/B). *Yersiniae* are psychrophilic, which means that they can not only survive, but proliferate at low temperatures such as 4℃.

Only two other members of this genus, *Yersinia enterocolitica* and *Yersinia pseudotuberculosis* cause infections in humans, whereas the others are thought to be environmental species (Sulakvelidze 2000). *Y. enterocolitica* and *Y. pseudotuberculosis* are transmitted via the fecal-oral route and cause gut-related symptoms such as diarrhoea (Wren 2003). These infections are relatively harmless for humans but especially *Y. pseudotuberculosis* can lead to devastating losses in animals, such as cattle (Tauxe 2004). Interestingly, *Y. pestis* is most closely related to *Y. pseudotuberculosis* and is thought to have developed from it 1,500 – 20,000 years ago (Achtman *et al.* 1999). This is surprising as *Y. pestis* causes a very different disease with a different infection route than *Y. pseudotuberculosis*. The main differences between these species are the inactivation of proteins required for entry via the intestine and acquisition of two additional virulence plasmids allowing transmission via fleas and systemic dissemination. This makes *Y. pseudotuberculosis* a good model for studying *Y. pestis* and bacterial pathogen evolution.

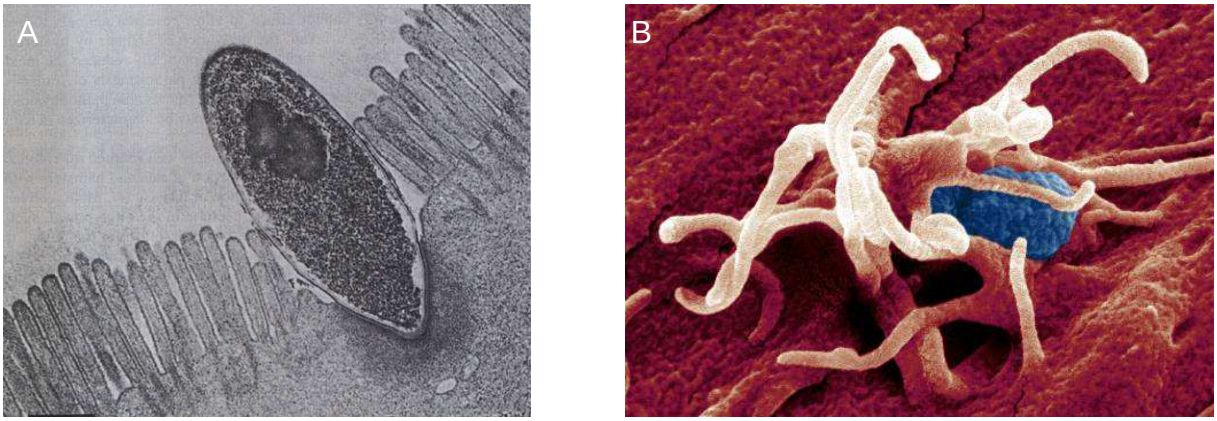


Fig. 1: Yersinia bacteria on host cells. A: *Yersinia pseudotuberculosis* YPIII on the surface of HEp-2 epithelial cell. With courtesy of Petra Dersch. B: *Yersinia enterocolitica* on the surface of HEp-2 epithelial cell. With courtesy of Frank Uliczka and Manfred Rohde.

1A.2 *Yersinia pestis*

Y. pestis is transmitted mainly via fleas and spreads with infected rodents such as rats (Fig. 2) (Wren 2003). Once inside the human body, the bacteria travel to the lymph nodes where they cause an inflammation leading to painful swellings called bubo (bubonic plague) (Zhou & Yang 2009). In some cases the disease can spread to the lung (secondary pneumonic plague) and can then be transmitted to other humans via droplet infection (leading to primary pneumonic plague). The development of pneumonic plague goes along with a severely decreased chance of survival (14% vs. 57% fatality rate of bubonic vs. pneumonic plague) (Inglesby *et al.* 2000). A minority of patients also develops a sepsis by *Y. pestis* without forming bubos (septicemic plague).

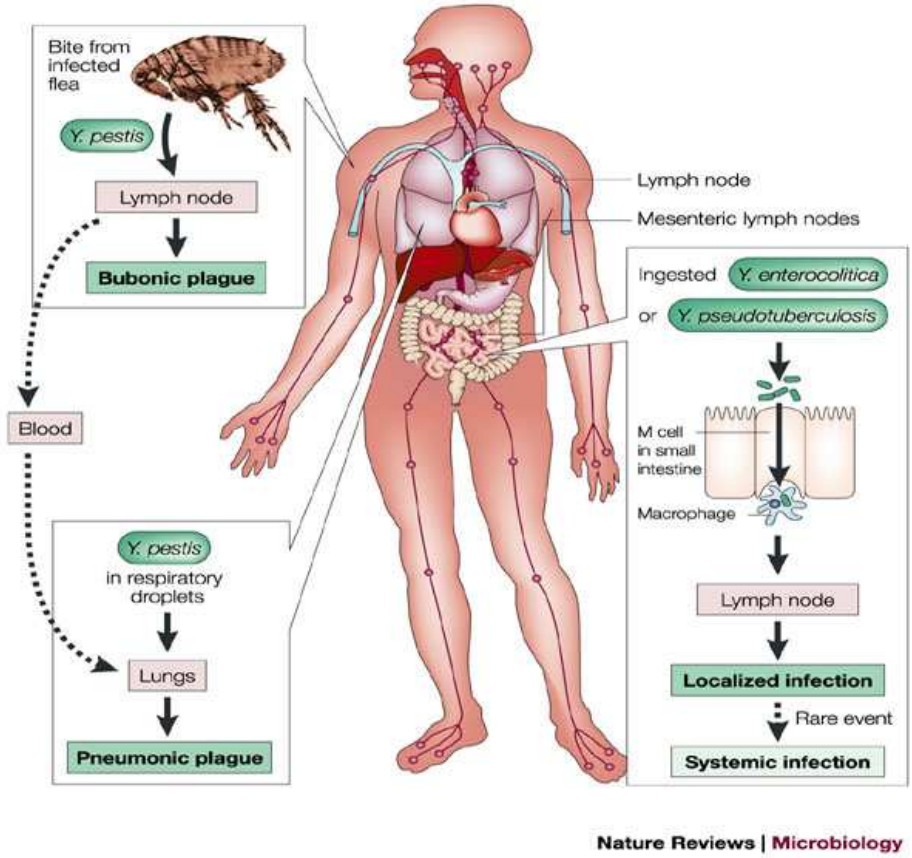


Fig. 2: Infection routes of pathogenic *Yersinia*. *Y. pestis* is transmitted by flea from animal reservoirs such as rats. The bacteria travel to the lymph nodes causing swelling due to the inflammation (bubonic plague). They can also travel to the lungs, where they establish the pneumonic plague which can be spread to other humans via droplet infection. The enteropathogenic *Yersiniae* are ingested with contaminated food or water and cross the intestinal barrier to infect the underlying lymph nodes. This leads to a localized infection causing gut-related symptoms such as enteritis or diarrhoea (adapted from(Wren 2003)).

1A.3 Enteropathogenic *Yersiniae*

The enteropathogenic *Yersinia* species *Y. pseudotuberculosis and Y. enterocolitica* are usually taken up by contaminated food, such as raw pork or water (Fig. 2) (Dube 2009). After passage into the intestine, the bacteria adhere to the mucosal surface, are then taken up by M-cells and are transported to the underlying lymphoid tissue, the Peyer´s patches (early stage of infection) (Heroven & Dersch 2010). From the Peyer´s patches the bacteria migrate to the liver and spleen, where they replicate extracellularly and lead to inflammation (late stage of infection) (Trulzsch *et al.* 2007). This entails the typical symptoms such as enteritis, mesenteric lymphadenitis and watery or bloody diarrhoea (Koornhof *et al.* 1999). Different sets of virulence factors are required for early and late phase infection, respectively (Fig. 3). Initial adhesion and invasion into cells are mediated by an outer membrane protein called invasin. It binds tightly to β_1- integrins, membrane receptors on the surface of human cells required for contact to extracellular matrix proteins such as laminin (Eble *et al.* 1998). In early phase infection and outside the host, the bacteria also need smooth lipopolysaccharide (LPS) and flagellar motility (Heroven & Dersch 2010). The early phase virulence factors are thus only produced at moderate temperatures. At higher temperatures in the host, these genes are repressed and the late phase virulence factors are expressed. Most of the late phase virulence genes are encoded on the Yersinia virulence plasmid pYV. They comprise the Yersinia adhesin A YadA (El Tahir & Skurnik 2001) and the components of the type III secretion system (Ysc, Yersinia secretion) as well as the corresponding effector proteins (Yop, Yersinia outer protein) (Heesemann *et al.* 2006; Matsumoto & Young 2009). The latter are injected into phagocytic cells and subvert several host cell components such as the actin skeleton to prevent phagocytosis (Trosky *et al.* 2008).

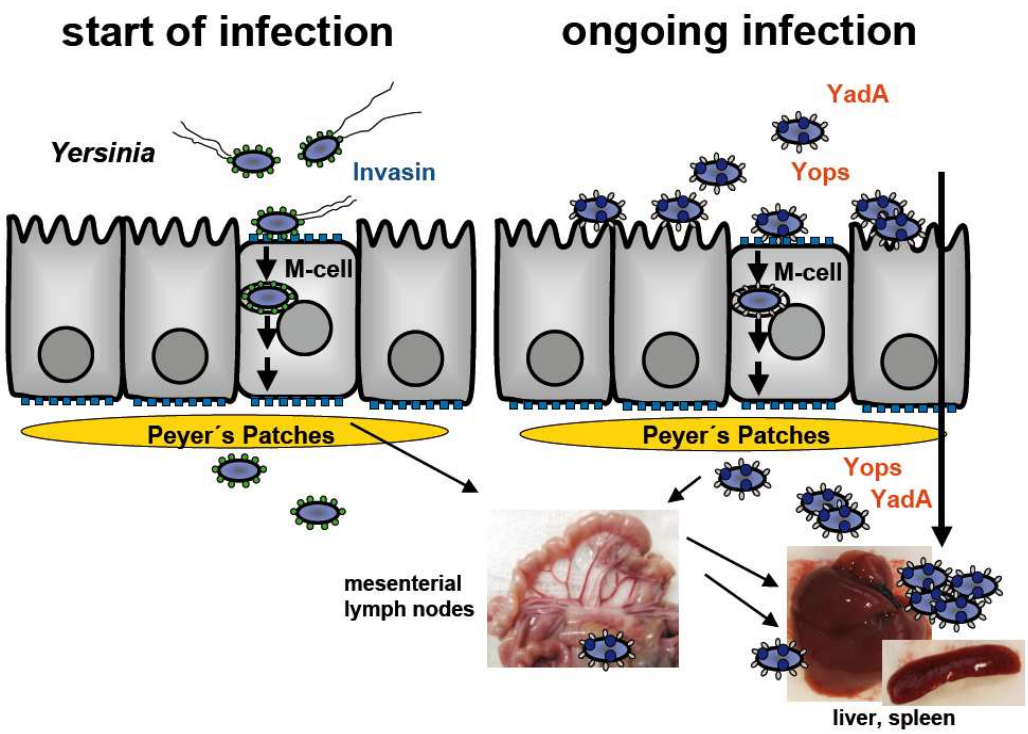


Fig. 3: Route of infection for enteropathogenic *Yersinia*. In the early phase of infection the bacteria use flagella and invasin to cross the intestinal barrier to reach the Peyer`s patches. During the later phases the bacteria reach the liver and spleen and use YadA for adhesion and the type III secretion system and the corresponding effectors to prevent phagocytosis (Heroven & Dersch 2010).

1A.4 RovA, a global virulence regulator

The expression of the early phase virulence factor invasin (*inv*) is tightly regulated in response to several environmental cues by a complicated regulatory network featuring several proteins and regulatory RNAs. The central node in this network where all the different influences are integrated is the protein RovA (regulator of virulence A) (Revell & Miller 2000). RovA belongs to the MarR-type family of transcription regulators that are involved in adaption to the environment, stress and virulence, with a central winged-helix DNA binding motif. RovA is mostly known to regulate invasin expression in enteropathogenic *Yersinia*, yet it also plays an important role in virulence regulation in *Y. pestis* which does not contain a functional *inv* gene. *rovA* mutants in *Y. pestis* have been shown to lead to improper membrane construction (Yang *et al.* 2010), reduction of pH 6 antigen expression (Cathelyn *et al.* 2006) and a significantly higher LD_{50} compared to the wild type (Ellison *et al.* 2004). Also, in enteropathogenic bacteria, RovA regulates many more genes than *inv*

(Cathelyn *et al.* 2007). This leads to a much more severe phenotype of *rovA* mutants compared to *inv* mutants (Nagel *et al.* 2001; Revell & Miller 2000).

1A.5 RovA regulation

The *rovA* gene in *Y. pseudotuberculosis* is expressed via two promoters P1 and P2 which are 76nt and 343nt upstream of the transcription start site. At 28℃ RovA binds to an AT-rich sequence upstream of P2 and thereby activates its own production (Heroven *et al.* 2004). There is an additional, low affinity binding site downstream of P1, which acts like an overpressure valve. RovA binding to this site leads to reduced transcription of *rovA* and prevents ever-increasing RovA production by its own positive feedback loop (Fig. 4). There is evidence that RovA might directly interact with the RNA polymerase to promote transcription of regulated genes (Tran *et al.* 2005), yet probably much more important is that RovA competes with the histone-like nucleoid structuring protein H-NS for binding at the *rovA* and *inv* promoter regions (Ellison & Miller 2006; Heroven *et al.* 2007). H-NS also binds to AT-rich sequences and is thought to silence gene expression by binding cooperatively to the DNA and either bridging separate DNA segments or inducing a loop in a single binding site (Dame *et al.* 2005; Noom *et al.* 2007). In the *rovA* as well as the *inv* promoter region, the binding sites of H-NS and RovA superpose and thus, counteracting H-NS repression seems to be the primary way RovA activates gene expression (Heroven *et al.* 2004; Tran *et al.* 2005).

1A.6 RovA is a protein thermometer

RovA, and thereby invasin production are dependent on several environmental conditions such as temperature, pH and growth phase of the bacteria (Nagel *et al.* 2001). RovA is only expressed at moderate temperatures (2-25℃), but not at 37℃. This is due to the unique feature of RovA to act as a protein thermometer (Herbst *et al.* 2009). At moderate temperatures RovA binds to its own promoter region and thus activates its own transcription. Yet, at 37℃, RovA undergoes a reversible conformational change, which strongly reduces its DNA binding capacity (Fig. 4). Additionally, the protein is now susceptible to proteolytic degradation by e.g. the Lon protease (Herbst *et al.* 2009). This leads to strongly reduced RovA and thus invasin production. Thus, after being taken up by the host (temperature shift to 37℃) and

penetration of the intestinal barrier, the early virulence genes (e.g. *inv*) are downregulated to prevent recognition by the immune system.

Additionally, this thermosensing mechanism is subject to the growth phase of the bacteria (Heroven & Dersch 2010). While RovA is highly unstable at 37℃ during exponential phase, it is stabilised at stationary phase. This growth phase effect might be due to the production of a small molecule during stationary phase which binds and stabilizes RovA (personal communication Petra Dersch).

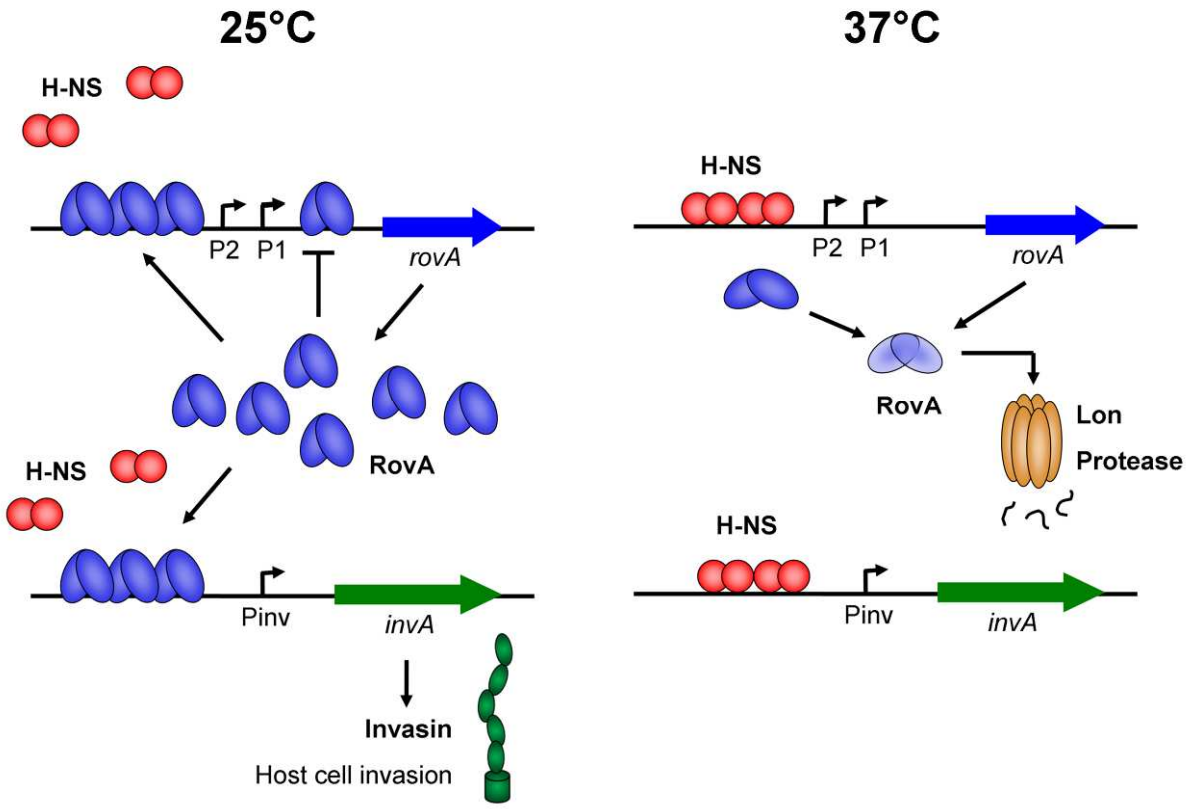


Fig. 4: Temperature regulated RovA expression. At 25℃ RovA binds to its own promoter region activating its own transcription by alleviating H-NS mediated repression. This allows production of invasin and therefore host invasion. At 37℃ the DNA binding capability of RovA is strongly reduced and it gets replaced by H-NS. Additionally, RovA is proteolytically digested by the Lon protease. This silences *inv* expression.

1A.7 RovM, an inhibitor of RovA

Another important component of this regulatory network is the LysR-type transcription regulator (LTTR) RovM (Heroven & Dersch 2006). LTTR proteins consist of two domains, an N-terminal winged helix-turn-helix (wHTH) DNA binding domain (DBD) and a C-terminal effector binding domain (EBD) linked by a long and flexible linker helix. Different oligomeric states have been observed for LTTR proteins: dimers (Zhou *et al.* 2010), tetramers (Monferrer *et al.* 2010; Muraoka *et al.* 2003), and octamers (Sainsbury *et al.* 2009). As a tetramer these proteins have two distant DBDs and are thought to bend the DNA around the protein (Deghmane *et al.* 2004). This leads to exposure of nucleotides in between, which then become accessible in a DNAseI protection assay (hypersensitive nucleotides). This has been shown to be also the case for RovM (Heroven & Dersch 2006), which binds to a region 50nt upstream of the *rovA* P1 promoter. In several cases it was observed that LTTRs downregulate their own expression (Schell 1993). In contrast, RovM seems to upregulate its own expression via an indirect mechanism (Heroven & Dersch 2006).

RovM acts together with H-NS to repress *rovA* transcription but in the absence of H-NS RovM is not able to repress *rovA* transcription. The main task for RovM seems to be the control of RovA (and thus early virulence genes such as *inv*) in response to nutrient availability (Heroven *et al.* 2007). Under conditions of low nutrient availability *rovM* is expressed, whereas *rovA* expression is repressed and vice versa. The signal for nutrient availability is relayed to RovM via the carbon storage regulator system (Csr) (Fig. 5) (Heroven *et al.* 2008). This system comprises one RNA binding protein CsrA and two regulatory RNAs CsrB and CsrC. The CsrA protein is responsible for activation of *rovM* expression by an indirect mechanism. The RNAs CsrB and CsrC bind several CsrA molecules and sequester them, preventing *rovM* activation. The production of CsrC is strongly regulated in response to growth phase and media composition, whereas the environmental signal for CsrB has yet to be determined. RovM has also been shown to be involved in regulating flagellar motility (Heroven & Dersch 2006).

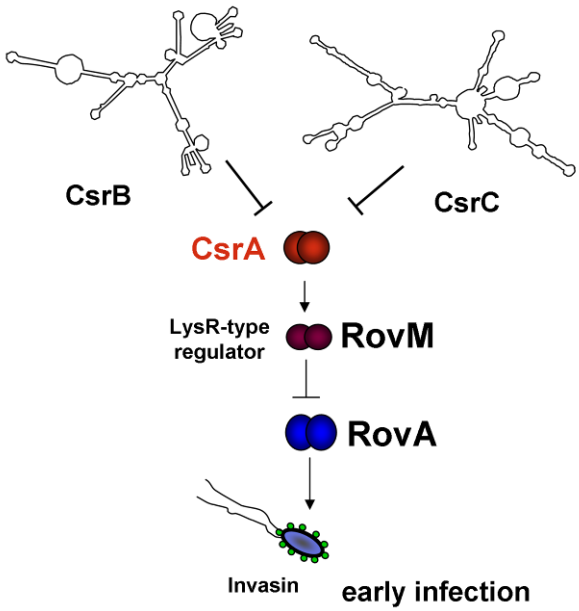


Fig. 5: Nutrient availability controls expression of early virulence genes via the carbon storage regulator (Csr) system and RovM. CsrB and CsrC bind to CsrA, sequestering it and preventing its activation of *rovM* expression. RovM can bind to the *rovA* promoter region in cooperation with H-NS to repress RovA and thereby Invasin production. With courtesy of Petra Dersch.

1B Polyketide synthases

Since ancient times, natural products have been used to treat illnesses. For example extracts from the bark of the willow tree were recognized to have an effect on fever and pain by the Greek physician Hippocrates. These findings led to the development of the drug acetylsalicylic acid, Aspirin, at the end of the 19th century. Another keystone in the development of modern drugs was the isolation of penicillin by Sir Alexander Fleming in 1928. With penicillin it was possible for the first time to effectively treat bacterial infections. Since then many different antibiotics have been discovered and developed and as a result bacterial infections nowadays do not pose such a serious threat as they used to. Yet, some bacteria are notoriously difficult to treat such as *Mycobacterium tuberculosis*, the causative agent of tuberculosis (Becerra *et al.* 2010). Especially worrying is the progressive spreading of multi-drug resistant bacteria strains, such as MRSA, methicillin-resitant *Staphylococcus aureus* (Boucher *et al.* 2010), and *Pseudomonas aeruginosa* (Hirsch & Tam 2010), both of which are virtually untreatable with common antibiotics and often leave the physicians with very few options. Thus, it is important to find novel antibiotics that are effective against multi-drug resistant bacteria.

1B.1 *Streptomycetes*

Most antibiotics originate from natural sources and are then chemically altered to enhance their properties such as efficacy. The most yielding source of antibiotics have been *Streptomyces* species (Wilkinson & Micklefield 2007). *Streptomycetes* are gram-negative bacteria from the order of Actinomycetales, which also comprises the pathogens *M. tuberculosis* and *C. diphtheriae* (Fig. 6). *Streptomycetes* are characterised by their very large, GC-rich genomes, containing more predicted genes than, for example, *Saccharomyces cerevisiae* (Bentley *et al.* 2002). They are mostly found in the soil and are responsible for the earthy smell of soil because of the production of a volatile substance called geosmin (Jiang *et al.* 2006). Over 500 species of *Streptomycetes* are known so far and a few of them can cause infections in humans (Quintana *et al.* 2008) and in plants (Lerat *et al.* 2009).

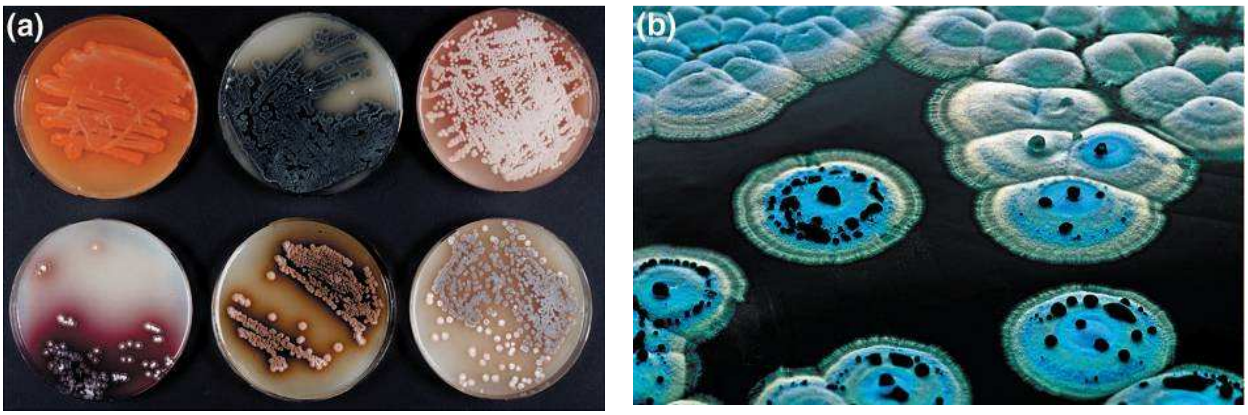


Fig. 6: *Streptomyces* colony morphology and pigmentation. A: *Streptomyces* colonies isolated from soil, producing colored pigments. B: *Streptomyces coelicolor* colonies with peripheral and aerial mycelia producing the blue antibiotic actinorhodin. (Thompson *et al.* 2002)

Yet, the most interesting feature of S*treptomycetes* is their complex secondary metabolism. Two thirds of all antibiotics used today, such as chloramphenicol, tetracycline and streptomycin have their origin in *streptomyces* bacteria (Fig. 7). These antibiotics are probably produced to defend their ecological niche against other bacteria (Thompson *et al.* 2002). *Streptomycetes* also produce other pharmaceutically valuable molecules such as immunomodulators (rapamycin) and anti-cancer drugs (doxorubicin). Therefore, these bacteria are continuously screened for novel antibiotics and other useful molecules which could be used for example as anti-cancer drugs (Wilkinson & Micklefield 2007).

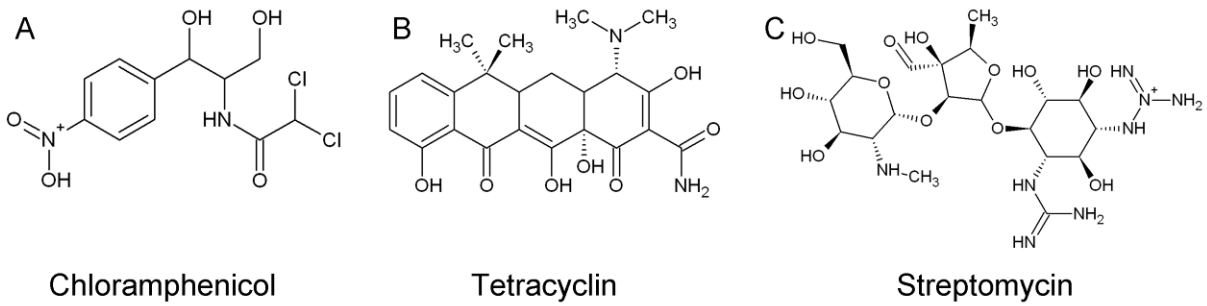


Fig. 7: Antibiotics produced by *Streptomycetes*: A: Chloramphenicol (*S. venezuelae*), B: Tetracyclin (*S. aureofaciens*), C: Streptomycin (*S. griseus*)

1B.2 Polyketide synthases

Most of these molecules are at least partially produced by so called polyketide synthases (PKS). PKSs are huge multi-enzyme complexes, that build up polyketides in a step-by-step manner, similar to an assembly line (Weissman & Muller 2008). They are organized in so-called modules, each responsible for the recognition and addition of one specific building block, called extender unit, to the growing chain (Khosla *et al.* 2009). The first module (loading module, LM) recognizes one specific building block (starter unit), usually a fatty acid-coenzyme A (CoA) ester and transfers it via an acyltransferase (AT) domain onto a acyl-carrier domain (ACP) (Fig. 8). The ACP domain is post-translationally modified by the addition of a phosphopantheteine (PPant) residue to a conserved serine by phosphopantheteine transferases (Lambalot *et al.* 1996). This PPant-“arm” is very flexible and able to reach the different domains of the protein and has a thiol group at its end (Lai *et al.* 2006). The starter unit is bound to this thiol group as a thioester by a trans-esterification reaction from the CoA region of the starter unit, which is then released.

The bound starter unit is then transferred to the next module which has already bound an extender unit (usually a carboxy- fatty acid-CoA ester) at its ACP domain. The starter unit from the first module is then added to the extender unit from the second module via a thioclaisen condensation in the ketosynthase (KS) domain of the second module, which sets free the “empty” ACP of the first module. The product is then transferred to the next module, where a new extender unit is added to the chain and so forth. During this whole process, the growing chain is covalently bound to an ACP domain of the PKS complex. This so called “substrate channelling” is important to ensure rapid production of the molecule and stabilisation of possibly instable intermediates (Spivey & Ovadi 1999). In the end, the polyketide is detached from the ACP of the last module by a thioesterase domain. This domain cleaves the thioester and often promotes cyclisation of the product (Akey *et al.* 2006).

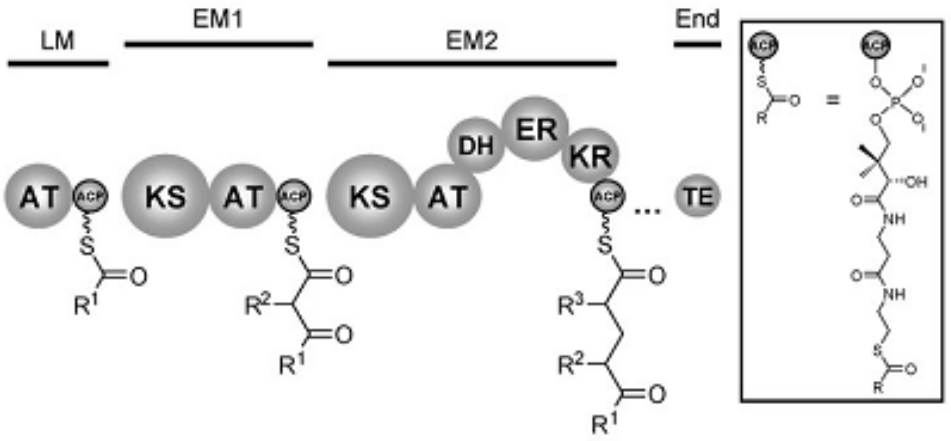


Fig. 8: Modular organization of PKS systems. A loading module (LM), consisting of a acyltransferase (AT) domain and an acyl carrier protein domain (ACP), binds the starter unit and transfers it to the first extension module (EM1). EM1 is a minimal module, containing also a ketosynthase (KS) domain, responsible for connecting the starter unit with the first extender unit. The second extension module (EM2) contains additional reductive domains (ketoreductase (KR), dehydrogenase (DH) and enoylreductase (ER), leading to a fully reduced methylene group instead of a keto group. A terminal thioesterase (TE) domain finally cleaves off the product from the protein and may catalyse cyclisation of the product. (Weissman & Muller 2008)

Thus, a typical PKS module consists of an AT, an ACP and a KS domain. In addition to which, there can be for example a ketoreductase (KR) domain, which reduces the keto group of the freshly added extender unit to an alcohol (Fig. 8). This can then be reduced to a double bond by a dehydrogenase (DH) domain and further to a fully reduced methylene group by an additional enoylreductase domain (ER). Additional tailoring can be done by methylase domains.

The reactions catalysed by PKSs are reminiscent of the related fatty acid synthases (FAS), yet the FAS can only do one set of reactions over and over to produce a uniform product (Maier *et al.* 2008). On the other hand, the modular architecture of PKS, where every module is capable of using a different extender unit, augmented by the variety of possible reductases and further tailoring allow PKSs to produce a huge variety of different molecules. The product of one PKS is often already obvious from the arrangement of the genes in the biosynthetic cluster (Challis *et al.* 2000). This makes PKS valuable targets for bioengineering efforts to change for example the arrangement of the modules and thus the structure of the product, allowing the production of altered versions of known drugs to enhance their properties to create completely novel molecules (Gokhale *et al.* 2007). Unfortunately, structural

information about PKS systems is still scarce, hampering rational engineering of these proteins.

1B.3 Cinnabaramides - potential new anti-cancer drugs

Extracts from *Streptomycetes* are continuously screened for interesting biological activities, such as antibiotic activity or anti-tumor activity. Recently, molecules, called cinnabaramides, were identified from a streptomyces extract that exhibit a potent inhibition of the human 20S proteasome (Fig. 9B) (Stadler *et al.* 2007). These molecules are very similar to salinosporamides from *Salinispora tropica*, which are currently under development as anti-cancer drugs (Fig. 9C) (Chauhan *et al.* 2005). The main difference is the chlorinated butyrate moiety in the salinosporamides is replaced by an octanoic residue in cinnabaramides. The gene cluster responsible for the production of cinnabaramides contains 19 predicted genes of which six (cinA-F) are involved in the cinnabaramide core biosynthesis (Rachid *et al.* 2011). CinA comprises two PKS modules, the first binds acetyl-CoA as a starter unit and the second elongates the chain with the unusual extender unit hexylmalonyl-CoA (Fig. 9A). CinA also contains one domain from a non-ribosomal peptide synthetase (NRPS) which work analogous to PKSs, but utilize amino acids as extender units to produce small polypeptides. Two other NRPS domains are located in CinB. These, together with the NRPS domain in CinA, are proposed to elongate the intermediate by a cyclohexenyl-alanine residue. The stand-alone KS domain CinC then catalyses the intramolecular condensation to generate the γ-lactam-ring. The thioesterase CinE cleaves off the thioester bound intermediate and promotes β-lacton-ring-formation. The predicted P450 monooxigenase CinD probably catalyses a hydroxylation to produce the β-hydroxy-2-cyclohexenyl-alanine residue in some cinnabaramides.

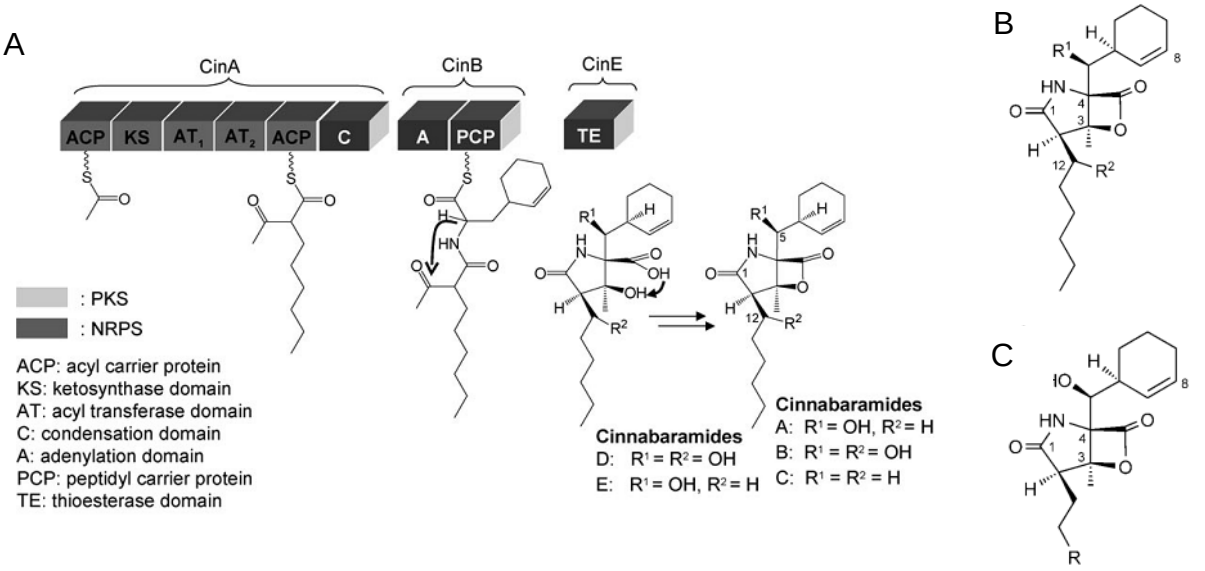


Fig. 9: Cinnabaramide synthesis. A: Scheme showing the biosynthesis of the Cinnabramides. B: Cinnabaramide A (R^1=OH,R^2=H), C: Salinosporamide A (R=Cl) and salinosporamide B (R=H) from *Salinispora tropica*. (Rachid *et al.* 2011)

1B.4 CinF produces an unusual PKS building block

The protein CinF shows similarity (50% sequence identity) to a crotonyl-CoA carboxylase/reductase (CCR) that generates ethylmalonyl-CoA by reductive carboxylation of crotonyl-CoA (Erb *et al.* 2007). Interestingly, these proteins use NADPH and CO_2 to catalyse this reaction (Erb *et al.* 2009). This is unique as usually carboxylation reactions require the presence of metals and often biotin and ATP as cofactors. Thus, it was proposed that CinF produces the unusual extender unit hexylmalonyl-CoA from octenoyl-CoA using NADPH and CO_2 (Fig. 10). Another protein that is proposed to catalyse this reaction is TgaD, a protein involved in the production of thuggacins in the myxobacterium *Sorangium cellulosum* (Buntin *et al.* 2010a). Surprisingly, the sequence identity is very low at only 15%. Carboxylases/reductases are very promising targets for bioengineering as changing their substrate specificity may allow the production of drugs featuring novel functionalities. Yet, such work requires structural information about these proteins and their reaction mechanism. This would then lead to the identification of the amino acids responsible for substrate specificity and allow rational design of new proteins with altered specificity. However, as yet, no structural information about this interesting class of proteins has been published.

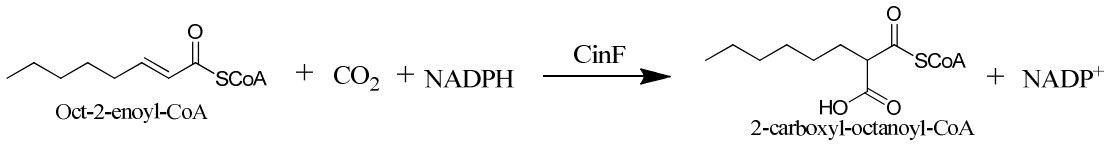


Fig. 10: Reaction catalysed by CinF. Octenoyl-CoA is converted to hexylmalonyl-CoA by CinF using CO_2 and NADPH as cofactors.

1B.5 Spirangienes

Spirangienes are highly cytotoxic and antifungal molecules, produced by the epothilon-producing myxobacteria strain *Sorangium cellulosum* So Ce90 (Fig. 11A) (Niggemann *et al.* 2005). They contain the characteristic spiroketal core structure which is also found in several other pharmaceutically important molecules such as avermectins (Albers-Schoenberg *et al.* 1981). The Spirangien biosynthesis cluster comprises 13 genes, coding for seven PKS proteins with a total of 15 modules for the production of the spirangien core structure (Frank *et al.* 2007). Six additional genes code for example for methyl transferases and monooxigenases for the tailoring of the spirangienes. Of special interest is the thioesterase domain of the last module 15 (SpiJ) called SpirTE (Fig. 11B). Thioesterase domains are responsible for the cleavage of the finished polyketide from the PKS. Most TE domains additionally catalyse the cyclisation of the cleaved product. SpirTE on the other hand releases a linear product without cyclisation. Understanding the molecular mechanism of TE domains is important as the TE domain has to be adapted to the new substrate in order to produce novel substances. Structural data is available for cyclising TE domains (Akey *et al.* 2006; Samel *et al.* 2006; Tsai *et al.* 2001), yet only recently one structure of a non-cyclising TE domain has been published without a substrate (Scaglione *et al.* 2010).

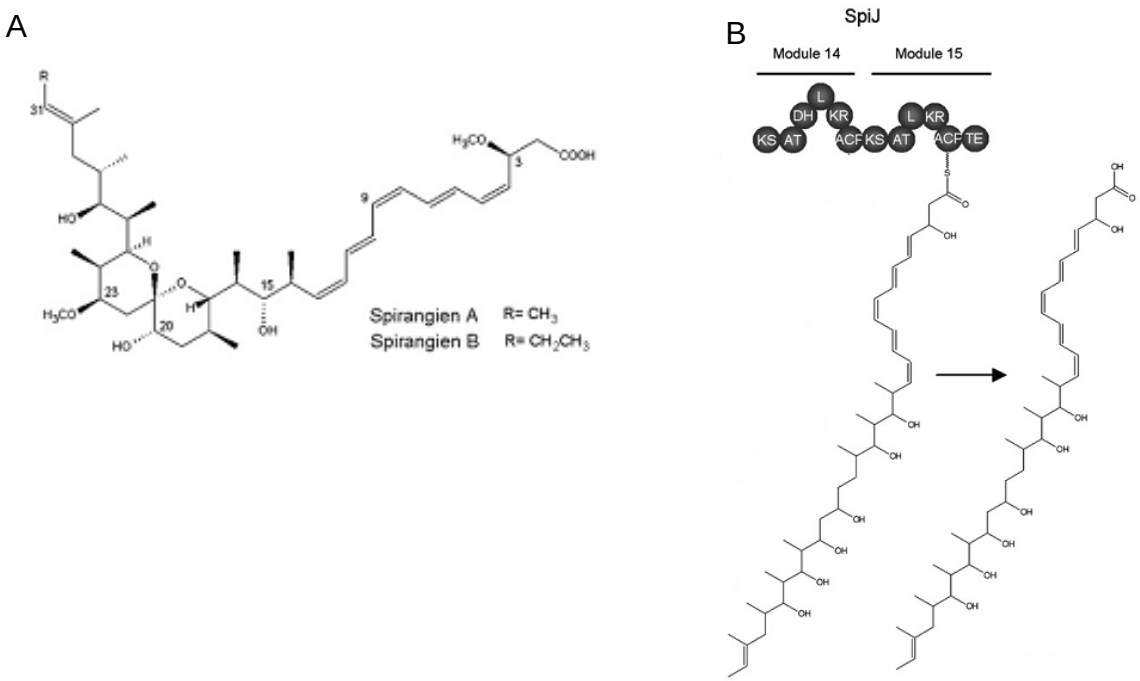


Fig. 11: Spirangien biosynthesis. A: Structure of Spirangiens A and B; B: Reaction catalysed by SpirTE in the Spirangien biosynthesis (modified from (Frank *et al.* 2007))

2 Aims

2A Virulence regulation in pathogenic *Yersiniae*

For efficient invasion into the host, enteropathogenic *Yersiniae* use the outer membrane protein invasin. Its expression is regulated by a complex network of regulatory proteins and RNAs in response to a large variety of environmental cues. The key protein in this network is RovA as it is directly responsible for invasin regulation in response to various signals such as temperature. The main aim of this thesis was to determine the structure of RovA alone and in complex with DNA to understand its DNA binding mechanism and the promiscuity it shows towards its DNA binding sequences. Additionally, the influence of small ligand molecules on RovA was to be elucidated structurally and biochemically. Furthermore, by comparison of RovA with its homologue SlyA from *S. typhimurium* the structural and functional basis of RovA thermosensing should be analysed.

The second aim was to determine the structure of RovM, an inhibitor of RovA. The structure should explain how RovM links the availability of nutrients, mediated by the CSR system, to virulence factor regulation by RovA.

2B Polyketide synthases

Engineering polyketide synthase systems promises the possibility of producing tailor-made molecules which can be used as drugs such as antibiotics. To be able to do this efficiently the molecular working mechanisms of these systems have to be understood in detail. Therefore, our first goal was to structurally and functionally characterise carboxylase/reductase enzymes, which provide the building blocks for polyketide synthases by a novel reaction mechanism. We aim to solve the first structure of one of these proteins, CinF, in complex with its substrates, to understand its unique reaction mechanism. Additionally, the structure should provide insights into the determinants of substrate specificity.

Furthermore, we wanted to solve the structure of SpirTE, a PKS thioesterase domain, to understand its reaction mechanism which leads to a linear product instead of a typically cyclised product.

3 Materials and methods

3.1 Materials

The chemicals that were used in this work were purchased from the following companies in the quality p.a. (*pro analysis*): Eurofins MWG Operon, Fermentas, Fluka, GE-Healthcare, Hampton Research, Invitrogen, Merck, Millipore, Qiagen, Riedel de Haen, Roche, Roth, Sigma-Aldrich and Stratagene.

3.1.1 Enzymes

Table 1: Enzymes and standards used in this work

Enzyme	Company
DNase I	Roche
T4 DNA Ligase	New England Biolabs
Platinum Pfx DNA Polymerase	Invitrogen
Pfu Turbo polymerase	Stratagene
Restriction endonucleases: NdeI, XhoI	New England Biolabs
Standards	
Smart Ladder (DNA)	Eurogentec
Precision Plus Protein All Blue Standards (Protein)	BIO-RAD

3.1.2 Kits

Table 2: Kits used in this work

Kit	Company
QIAprep® Spin Miniprep Kit	Qiagen
QIAquick® PCR Purification Kit	Qiagen
QuikChange® Site-Directed Mutagenesis Kit	Stratagene

3.1.3 Screens

The following crystallization screens obtained from Qiagen were used in this work:

Table 3: Crystallization screens used for finding initial crystallization conditions

The AmSO4	The pHClear
The Anions	The PACT
The Cations	ProComplex
The Classics I + II	JCSG+
The Cryo	JCSG Core I-IV
The PEGs I + II	

The Morpheus screen from Molecular Dimensions was also used, as well as the Additive Screen from Hampton Research.

3.1.4 Oligonucleotides

In the following DNA oligonucleotides (primers) are listed that were used in this work. These were purchased from Eurofins MWG Operon. The restriction sites are underlined and the corresponding restriction enzyme listed. For primers used for mutagenesis the desired mutation in the protein is given.

Table 4: List of DNA oligonucleotides

Primer	Sequence (5`→ 3`)	Restriction enzyme
Primers used for cloning		
RovM EBD FW	gggaattc**catatg**tacagcaatatggaagg	NdeI
RovM EBD RV	aggaga**ctcgag**gactacttaatc	XhoI
Primers used for mutagenesis		
RovA G6F FW	ccatggaatcgacattattctctgatctagcacgattag	
RV	ctaatcgtgctagatcagagaataatgtcgattccatgg	
RovA A10F FW	cattaggatctgatctattccgattagttcgcgtttggc	
RV	gccaaacgcgaactaatcggaatagatcagatcctaatg	
Oligonucleotides used for determination of minimal RovA binding fragment		
29 FW	taattgatattatttatatgataatagtt	
RV	aactattatcatataaataatatcaatta	
27 FW	aattgatattatttatatgataatagt	
RV	actattatcatataaataatatcaatt	
25 FW	attgatattatttatatgataatag	
RV	ctattatcatataaataatatcaat	
23 FW	ttgatattatttatatgataata	
RV	tattatcatataaataatatcaa	
21 FW	tgatattatttatatgataat	
RV	attatcatataaataatatca	
19 FW	gatattatttatatgataa	
RV	ttatcatataaataatatc	
17 FW	atattatttatatgata	
RV	tatcatataaataatat	
15 FW	tattatttatatgat	
RV	atcatataaataata	
13 FW	attatttatatga	
RV	tcatataaataat	
11 FW	ttatttatatg	
RV	catataaataa	
20a FW	gatattatttatatgataat	

	RV	attatcatataaataatatc
20b	FW	tgatattatttatatgataa
	RV	ttatcatataaataatatca
19a	FW	atattatttatatgataat
	RV	attatcatataaataatat
19b	FW	tgatattatttatatgata
	RV	tatcatataaataatatca

Oligonucleotides used for cocrystallization			Overhangs
Inv1 20	FW	tgatattatttatatgataa	Blunt end
	RV	ttatcatataaataatatca	Blunt end
Inv1 21	FW	ttgatattatttatatgataa	Blunt end
	RV	ttatcatataaataatatcaa	Blunt end
Inv1 20A	FW	atgatattatttatatgataa	5`A
	RV	tttatcatataaataatatca	5`T
Inv1 25A	FW	aaattgatattatttatatgataata	5`A
	RV	ttattatcatataaataatatcaatt	5`T
Inv1 19C	FW	cgatattatttatatgatc	5`C
	RV	cgatcatataaataatatc	5`C
Inv1 19G	FW	ggatattatttatatgatc	5`G
	RV	ggatcatataaataatatc	5`G
Inv1 Art	FW	atatgataaatgatattattt	3`A
	RV	aataatatcatttatcatata	3`A
Inv2a	FW	aatagactgttttattatata	5`A
	RV	ttatataataaaacagtctat	5`T
Inv2b	FW	aataaaacagtctataccata	5`A
	RV	ttatggtatagactgttttat	5`T
rovA1A	FW	aattatattatttgaattaat	5`A
	RV	tattaattcaaataatataat	5`T
rovA1B	FW	tttacatccattaattatata	5`A
	RV	atatataattaatggatgtaa	5`T
rovA2A	FW	tatgctagcacgctaattaaa	5`A
	RV	atatgctagcacgctaattaa	5`T

3.1.5 Plasmids

Table 5: Plasmids used in this work

Plasmid	GOI	Derivation	Reference
pET28-RovA	RovA C81S, C108S	Site directed mutagenesis of *rovA* in pET28a	This work
pET28-RovA G6F	RovA C81S, C108S, G6F	Site directed mutagenesis of pET28-RovA	This work
pET28-RovA A10F	RovA C81S, C108S, A10F	Site directed mutagenesis of pET28-RovA	This work
pAKH43	RovM	*rovM* in pET28a	Heroven et al 2006
pET28-RovM-EBD	RovM-EBD	*rovM*-EBD (aa 92-310) in pET28a	This work
pET28-SpirTE	SpirTE	*spirTE* in pET28a	Buntin et al. 2010

3.1.6 Bacterial strains

Table 6: Bacterials strains

Strain	Genotype	Company
Top10	F^{-}, mcr A Δ(mrr-hsdRMS-mcrBC) Φ80 lacZ Δ M15 ΔlacX74 recA1 araD139 Δ(ara-leu)7697 gal U gal K rpsL endA1nupG	Invitrogen
XL1-Blue	recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F´proAB lacIqZΔM15 Tn10 (Tetr)].	Stratagen
BL21	F^{-} ompT hsdSB(rB^{-} mB^{-}) dcm^{+} dam^{+} Tet^{R} galλ (DE3) endA Hte [argU, ileY, leuW, Cm^{R}]	Stratagen

3.1.7 Media and buffers

Table 7: Media and buffer compositions

Name	Composition
Media	
Luria Bertani (LB)	10 g/L tryptone, 5 g/L, NaCl, 5 g/L yeast extract
SOC-Medium	20 g/L tryptone, 5 g/L yeast extract, 0.5 g/L NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 20 mM glucose
10x M9	60 g/L Na_2HPO_4, 30 g/L KH_2PO_4, 5 g/L NaCl, 10 g/L NH_4Cl, autoclave separately
Minimal Medium (M9)	100 ml 10x M9, 0.1 mM $CaCl_2$, 1 mM $MgSO_4$, 20 mM glucose, 0.2 % casaminoacids
Buffers	
Bandshift buffer (10%)	100 mM Tris-HCl pH 7.5, 10 mM EDTA, 50 mM DTT, 50 % Glycerol, 100 mM NaCl, 10 mM $MgCl_2$, 1 mg/ml BSA
CD buffer	10 mM NaH_2PO_4, pH 8, 10mM NaCl, 5mM DTT, 1mM $MgCl_2$
DNA loading buffer (10x)	10 mM Tris/HCl, pH 7.5, 0.05% (w/v) bromphenol blue, 1 mM EDTA, 50% glycerol
Gel filtration buffer	20mM Tris/HCl pH 8, 100mM NaCl, 5mM DTT
Lysis buffer	50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 10 mM imidazole, 1 mM $MgCl_2$
Ni-NTA wash buffer	50mM NaH_2PO_4 pH 8.0, 300mM NaCl, 20mM imidazole
Ni-NTA elution buffer	50mM NaH_2PO_4 pH 8.0, 300mM NaCl, 250mM imidazole
RovA Gel filtration buffer	20mM Tris/HCl pH 9, 100mM NaCl, 5mM DTT
SDS-PAGE Lower buffer (4x)	1.5 M Tris/HCl, pH 8.8
SDS-PAGE running buffer	25 mM Tris/HCl, 192 mM glycine, 0.1% (w/v) SDS
SDS-PAGE sample buffer (2x)	1 ml 1 M Tris/HCl pH 6.8, 2.4 mM glycerol, 0.8 g SDS, 2 mg Coomassie blue G-250, 0.31 g DTT, add H_2O to 10 mL

SDS-PAGE separating gel	10 mL Polyacrylamide solution, 5 mL 4 x Lower buffer, 0.2 mL SDS (10%), 4.7 mL H_2O, 20 µL TEMED, 50 µL APS (25%)
SDS-PAGE stacking gel	1.5 mL Polyacrylamide solution, 2.5 mL 4 x Upper buffer, 5.9 mL H_2O, 15 µL TEMED, 25 µL APS (25%)
SDS-PAGE Upper buffer (4x)	0.5 M Tris/HCl pH 6.8, 0.4% (w/v) SDS
TAE buffer	40 mM Tris/HCl pH 8.2, 20 mM sodium acetate, 1 mM EDTA

3.2 Methods

3.2.1 Standard procedures

Standard methods used in this work are adapted from collections of methods and protocols (Ausubel *et al.* 2007; Coligan *et al.* 2002). Only variations of standard protocols are described below.

3.2.2 Protein production and purification

All proteins in this work were cloned into a pET28 vector and were thus produced and purified in a very similar way. A standard protocol is given here and exceptions are mentioned where appropriate.

The vector was transformed into competent *E. coli* BL21 (DE3) cells and single colonies were selected on a Kanamycin-containing (30µg/ml) LB-plate. One colony was used to inoculate a 50ml overnight LB culture containing Kanamycin (30µg/ml). The next day 2x1L of LB (M9 for RovA) medium were inoculated with 10ml of the overnight culture each and grown at 37℃ at 160rpm. At an OD_{600} of 0.7 the cultures were induced with 0.25mM of IPTG and incubated at 20℃ over night. Then the cells were harvested by centrifugation at 6,000 rpm at 4°C for 10min. The supernatant was discarded and the pellet was stored at -20℃.

3.2.2.1 Production of SeMet-labelled protein

For the production of Se-Met labeled protein, an overnight culture was grown in LB medium as described above. The cells centrifuged at 5.000 rpm for 15min, washed in water and centrifuged again. The supernatant was discarded and the cells resuspended in minimal medium to an OD_{600} of 1.0. Selenomethionine was added to a final concentration of 15µg/ml medium as well as 0.25mM IPTG to induce protein production. The cells were incubated over night at 20℃ and subsequently harvested by centrifugation.

3.2.2.2 Protein purification

The cell pellet was resuspended in 50ml Lysis buffer and DNaseI and a Complete Protease Inhibitor Cocktail Tablet (Roche) were added. The cells were lysed two times by using a homogenizer at 20kpsi at 4℃. The lysate was cleared from cell

debris by centrifugation at 16,000rpm for 45min at 4℃. The supernatant was applied to a Ni-NTA column pre-equilibrated with Wash buffer and incubated under mild shaking at 4℃ for 30min. Unspecifically bound protein was removed by washing the column with 3 column volumes (about 50ml each) of Washing buffer. The target protein was eluted with 4x 15ml Elution buffer. The eluted fractions were tested for protein yield and purity by SDS PAGE (15% for RovA, 12% for the other proteins).

The chosen protein fractions were pooled and concentrated to 1ml using Vivaspin concentrators with a molecular weight cutoff (MWCO) of 10kDa. For further purification and removal of the imidazole the protein was applied to a 16/60 Superdex200 (SD200) gel filtration column pre-equilibrated with Gel filtration buffer (for RovA: RovA Gel filtration buffer). The protein containing elution fractions were again tested by SDS PAGE and the fraction containing the pure protein were pooled and concentrated to 10mg/ml (50mg/ml for RovA).

3.2.3 Protein analytical methods

3.2.3.1 SDS polyacrylamide gelelectrophoresis (SDS PAGE)

SDS PAGE was used to evaluate the purity and yield during the protein production (Laemmli 1970). The gels were composed of two parts the upper stacking gel (5% acrylamide) and the lower resolving gel (12% or 15% acrylamide). Typically 6µl of protein solution together with 6µl of SDS loading buffer were heated for 5min at 90℃ and then run at constant 40mA per gel on an SDS gel until the dye front reached the end of the gel (usually 30min). After the electrophoresis, the gel was washed with water to remove SDS and then stained with Instant Blue coomassie dye for 5min.

3.2.3.2 Mass spectrometry (MS)

Mass spectrometry (MS) analysis was performed to confirm the identity and the correct labeling of the produced proteins using either MALDI-TOF (Matrix Assisted Laser Desorption Ionisation – Time Of Flight) or ESI-MS (Electrospray Ionisation). These analyses were carried out by Dr. Manfred Nimtz and Undine Felgenträger (HZI, Braunschweig).

3.2.3.3 Gel shift assay

Gel shift assays are used to determine the affinity of DNA binding proteins to DNA fragments. The DNA is submitted on a native polyacrylamide gel electrophoresis with and without protein, where the protein bound DNA is retarded in respect to the free DNA due to the increased size.

Before the experiment the single stranded oligo nucleotides were mixed in an equimolar ration, heated to 95℃ for 5min and then let cool slowly to allow double strand formation. In our studies 150ng of DNA were used together with 7.5µg RovA in bandshift buffer and run on an 8% gel at 70V for 40min. Afterwards, the gel was stained in ethidium bromide solution (50µl/50 ml) and inspected under UV light.

3.2.3.4 ThermoFluor assay

ThermoFluor assays allow rapid screening of the influence of buffer conditions and additives on the stability of a protein. It is often used to find a buffer in which the protein is most stable. It can also be used to determine the binding affinity of a substance to a protein (Saridakis *et al.* 2008). The protein is mixed with a fluorescent dye and heated in 0.5℃ steps while the dye fluorescence is recorded. When the protein starts to unfold, hydrophobic patches become accessible to the dye. The dye is thereby shielded from quenching in the aqueous solution and the fluorescence increases. After the melting point protein aggregation reduces the available hydrophobic patches and the fluorescence drops slowly. From the resultant curve the melting point can be obtained by determining the peak position of the first derivative of the fluorescence curve. Comparison of the melting points for different buffer conditions reveals conditions in which the protein is most stable.

For testing of a set of standard buffers a screen similar to that suggested by Niesen et al., 2007 was used. The protein (1-2mg/ml end concentration) was diluted in the desired buffers and a fluorescent dye (Sypro Orange, 100x) was added. The samples were measured in a CFX96 Real-Time Thermal Cycler C1000 (BioRad) in a 96-well plate (Multiplate 96-Well Unskirted PCR plate, BioRad) sealed with a non-fluorescence absorbing seal (Microseal 'B' Adhesive Seal, BioRad). The corresponding CFX software as well as Microsoft Excel were used to analyse the data.

3.2.3.5 Circular dichroism (CD)

Circular dichroism is a method for the determination of the secondary structure contents of a protein as well as the protein stability. It is based on the differential absorption of left and right circularly polarized light by chiral molecules (such as proteins). Proteins are usually measured in the UV range and exhibit distinct absorption characteristics depending on their content of secondary structure elements.

Before the measurement the protein was diluted to 0.16mg/ml with CD buffer. Spectra were recorded from 190nm to 250nm at different temperatures with a J-815 CD spectrometer (Jasco). For denaturation curves the ellipticity at 208nm and 222nm (Minima for alpha helical proteins) was measured with increasing temperatures. Evaluation was similar to the melting curves from the thermofluor experiments.

3.2.4 Protein crystallization

3.2.4.1 Initial screening

To screen for initial crystallization conditions 0.2µl of the concentrated, pure protein (typically 10mg/ml) was mixed with the same volume of reservoir buffer of commercially available screens in a 96-well sitting drop vapor diffusion plate using the Honeybee 961 crystallization robot (Zinsser Analytics). The plate is sealed using a MancoTM Crystal Clear tape (Jena Bioscience) and the drops left to equilibrate against the reservoir solution at 16℃. After three days and the every two days the plates were checked for the appearance of crystals with a microscope.

3.2.4.2 Optimization

Initial crystals were tested for diffraction if possible and further optimized using 24-well hanging drop vapour diffusion plates. For the optimization several parameters such as the protein and precipitant concentration as well as the pH, different drop sizes and incubation temperature were varied. Additionally, crystal growth promoting additives were tested using the Additive Screen (Hampton Research) and seeding techniques such as micro-seeding and streak seeding (for RovA with salicylate, SpirTE and CinF) were employed to obtain larger crystals.

3.2.4.3 Cryo protection

X-rays induce the formation of radicals within the crystal during the measurement (Garman 2010). These can migrate through the crystal and damage it leading to rapid loss of diffraction. This is strongly reduced by freezing the crystals to 100°K (in liquid nitrogen). Yet, freezing of aqueous solutions results in ice formation which can damage the crystal and reduce the quality of the measured data by its own diffraction (ice rings). This is avoided by the addition of cryo protectants such as glycerol, small molecular weight PEG or sugars. For most crystals in this work a solution of reservoir buffer including 10-20% glycerol was prepared, the crystal transferred from its drop into this solution and then quickly frozen in liquid nitrogen. CinF was crystallized in a solution already containing glycerol and thus no further cryo protectant was required.

3.2.5 X-ray structure analysis

3.2.5.1 Diffraction of crystals

X-ray crystallography allows the determination of protein structures at atomic resolution. These structures can be used to elucidate the molecular mechanism of enzymes, design of specific protein inhibitors and understand the molecular details of protein interactions. For this a single protein crystal of sufficient size is required. This crystal is frozen and irradiated with monochromatic X-rays, generated either by a rotating anode X-ray generator (homesource) or a synchrotron. While most of the rays do not interact with the crystal, a small percentage is diffracted in a certain direction with certain intensity and these diffracted rays are measured on a detector (typically a CCD device). The crystal is rotated around one axis in small steps (0.5°) and one image is recorded for each step. From these recorded images the position and the intensity of the diffracted X-rays can be calculated providing information about the structure of the protein in the crystal.

In which direction and under which angle an X-ray is diffracted is determined by Bragg`s Law,

$$n\lambda = 2d\sin(\theta), \qquad \text{(equation 1)}$$

where *n* is an integer, *λ* is the wavelength of incident wave, *d* is the spacing between the planes in the atomic lattice, and *θ* is the angle between the incident ray and the scattering planes.

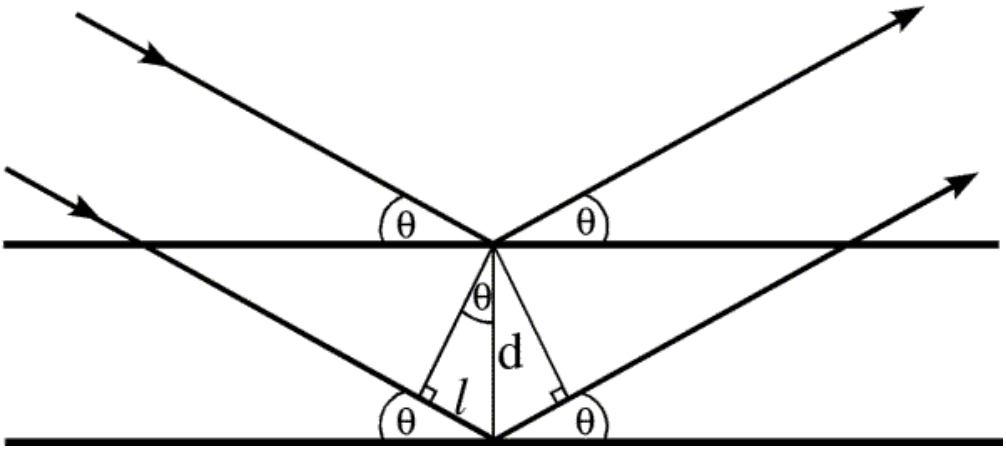


Fig. 12: Diffraction of X-rays from two parallel planes. To produce constructive interference the two reflected rays have to be perfectly in phase. This is only true when the distance the second ray has to travel more than the first (2*l=2dsinθ) is an integer multiple of the wavelength.

It assumes the presence of sets of (imaginary) planes (called Miller Planes, characterized by its three Miller coordinates *h*, *k* and *l*) which are evenly distributed throughout the crystal lattice that act as semitransparent mirrors (Fig. 12). Thus if an X-ray impinges on one plane from one set it will be reflected by the plane while another ray from the same direction will reach the plane below and be reflected by it. These two emerging waves will interfere with each other. This interference can only be constructive if the difference in the distances these two rays have to travel is equal to an integer number times the wavelength. As these sets of planes are considered to be infinite even a small phase difference between these two rays will lead to destructive interference with a wave coming from a plane much further away. Thus for each orientation of the crystal a certain number of planes will fulfill Bragg`s Law and produce a measurable diffracted X-ray. From the pattern that is generated by these rays, the crystal parameter and symmetry can be deduced.

3.2.5.2 Calculating the electron density

With the knowledge of the intensity and phase for each reflection, the electron density for the unit cell of the crystal can be calculated.

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}\left|F_{hkl}\right|e^{-2\pi\cdot i(hx+ky+lz-\alpha_{hkl})}, \qquad \text{(equation 2)}$$

where $\rho(x,y,z)$ is the electron density, V is the unit cell volume, $|F_{hkl}|$ is the structure factor amplitude for each reflection (which is equal to the square of the measured intensity) and α_{hkl} are the phases for each reflection.

Unfortunately, the phase information, which carries more information than the intensity, is lost during the measurement. As it is impossible to determine the phase directly, several indirect methods have been developed to determine the phase (or produce a good estimate of it). These methods include single/multiple isomorphous replacement (SIR/MIR), molecular replacement (MR), single/multiple wavelength anomalous dispersion (SAD/MAD) or combinations thereof. Only SAD and MR will be discussed further as these were used in this work.

3.2.5.3 Molecular replacement (MR)

Molecular replacement is an easy and often quick way to solve the phase problem by using phase information obtained from structures of other proteins with a similar fold. Obviously, this method suffers from the limitation that a suitable molecular replacement model has to be available. Yet, with the quickly growing number of protein structures in the Protein Database (PDB) this method becomes ever more powerful.

By applying the Fourier back-transformation to an already solved protein structure (and thus doing an imaginary diffraction experiment) the phase information for this protein can be obtained. Molecular replacement is based on the assumption that these phases of a homologous protein will serve as a sufficiently good estimate of the real phases. By combination with the measured intensity interpretable density can be obtained and further be refined. The main problem in this method is the generation of a good MR model and the orientation and positioning of the model. For MR to work, the model has to be oriented in the same unit cell in the same way as the unknown structure. A brute force approach is computationally very intensive as six parameters

(the three position coordinates and the three angles) have to be varied at the same time. To accelerate this process, the search for the correct orientation is performed first and with the found solution a positional search is done. This reduces one six-dimensional problem to two three-dimensional problem, which is much less complex. The search for the correct orientation (rotation search) uses the correlation between the Patterson maps from the known and the desired structure. The Patterson maps are generated by the application of the Patterson function, which is very similar to the equation 2 used to calculate the electron density, but does not include the phase information:

$$P(u,v,w) = \frac{1}{V} \sum_h \sum_k \sum_l \left| F^2{}_{hkl} \right| e^{-2\pi \cdot i(hu+kv+lw)} \qquad \text{(equation 3)}$$

The Patterson maps do not provide information about the absolute position of the atoms, but rather the relative distances between the different atoms. This information can be enough to solve the phase problem by trial and error for small molecules, but not for proteins. Yet, the Patterson map provides information about the orientation of the molecule. The Patterson map of the model is calculated for every possible orientation and the correlation with the Patterson map of the measured protein is calculated. The orientation that correlates best should be the orientation of the measured protein.

After the correct orientation has been found, the search for the position of the molecule is started (translation search). This search also relies on the calculation of Patterson maps with the correctly oriented model in different positions within the unit cell. The different symmetry elements are important in this search. Is the desired protein in the space group P1, which does not contain any crystallographic symmetry, this search is not necessary, as the unit cell can be defined as desired. Once the model is in the same orientation and position as the measured protein, the Fourier back-transform of the molecule is performed. This provides calculated phases and intensities. The intensities are discarded and replaced by the measured intensities. Together with the phases, the Fourier transform is calculated and a (hopefully) interpretable electron density is obtained.

3.2.5.4 Single/multiple wavelength anomalous dispersion (SAD/MAD)

Another popular technique to obtain phase information is anomalous dispersion. It relies on differences between Friedel mates (F_{hkl} and $F_{-h,-k,-l}$) that occur when wavelengths are used for the diffraction experiment, which are close to an absorption edge of an element incorporated in the crystal. Sometimes these elements are already present in the protein such as sulfur in cysteines and methionines, prosthetic groups such as iron sulfur clusters or bound metal ions. Alternatively, they can be introduced during protein production (SeMet labeling) or by soaking of a native crystal with heavy atoms such as samarium, iodine or gold. Anomalous dispersion is a powerful method as a single crystal is often enough to solve the structure and especially SeMet labeling can accelerate producing crystals containing suitable elements. The method does not require any prior knowledge about the structure of the measured protein and can thus be applied to every project, even if no suitable MR model is available and it avoids model bias.

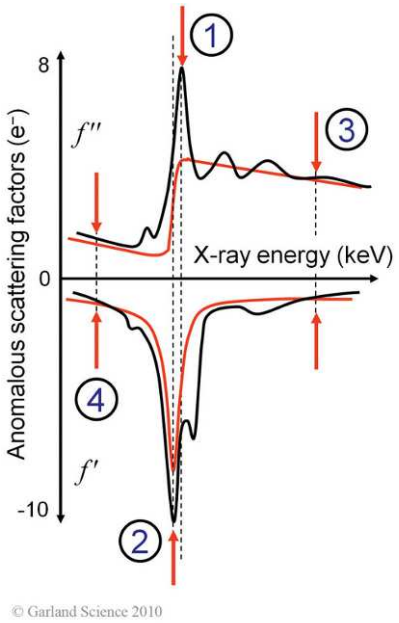


Fig. 13: The anomalous scattering factor contributions f´ and f´´ around an absorption edge. Red: theoretical curves; black: real values. The arrows mark wavelength commonly used for anomalous dispersion experiments. 1: The peak dataset contains the maximal anomalous differences (f´´ has its maximum value and f´ has a large negative value) and is thus usually the first dataset measured in a MAD experiment and the only one in a SAD experiment. 2: The inflection dataset carries the largest negative value for f´. 3: The high energy remote still has a high value for f´´, but nearly no f´. 4: The low energy remote carries no f´ and f´´ signal and is only of use in combination with the other datasets. (C) 2011 From Biomolecular Crystallography by Bernhard Rupp. Reproduced by permission of Garland Science/Taylor & Francis LLC.

Anomalous dispersion occurs near the absorption edges of elements within the crystals. The absorption of the X-rays leads to a small change (dispersive difference) in the intensity and phase of each reflection, described by the anomalous scattering factor contributions f´ and f´´, which are added to the structure factor of the anomalous scatterer (equation 4). These contributions are strongly wavelength dependent, f´ is minimal at the inflection point of the absorption edge, while f´´ is maximal at its peak (Fig. 13).

$$f_{(S,\lambda)} = f^{0}_{(S)} + f'_{(\lambda)} + i \cdot f''_{(\lambda)}, \qquad \text{(equation 4)}$$

with $f_{(S,\lambda)}$ the atomic scattering factor of the heavy atom with anomalous scattering and $f^{0}_{(S)}$ the scattering factor of the heavy atom without anomalous scattering.

The most important consequence of this is the breakdown of Friedels Law. Friedels Law states that two structure factors F_{hkl} and $F_{-h,-k,-l}$ which are a so called Friedel pair are complex conjugates of each other, meaning that their magnitude is the same, but their phases are inverse. In the presence of anomalous scattering F_{hkl} and $F_{-h,-k,-l}$ have different magnitudes and also their phases are no longer related. These differences can be exploited to calculate the structure factors for the anomalous scatterers alone, which allows the determination of the “structure” of the anomalous scatterers by direct methods. This also provides phase information of the structure factors of the anomalous scatterers. By combination with the measured structure factors two possible phases for each protein structure factor can be calculated. As the scattering factors f´ and f´´ are dependent on the wavelength, measurement at a different wavelength provides the information necessary to determine the correct choice between the two phases (MAD). An alternative approach is to combine the two possible solutions into a best guess as to what the phase may be and use complementary methods such as density modification to improve this guess (SAD) (Fig. 14).

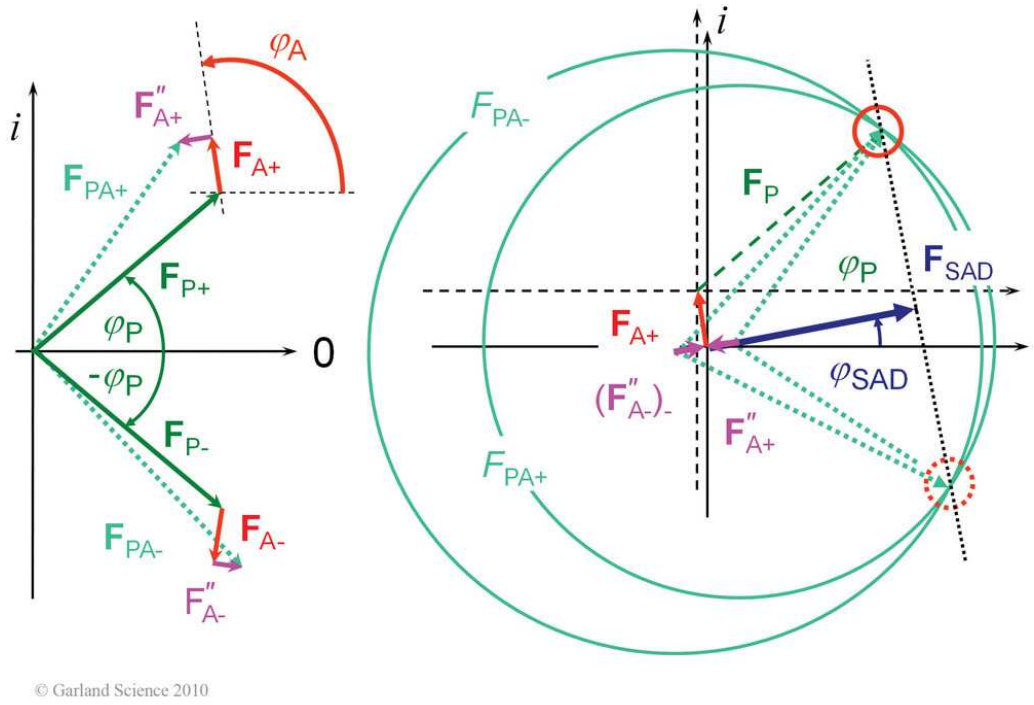


Fig. 14: Obtaining phase information from SAD data. Argand diagram of a protein structure factor (F_{P+}) and its Friedel mate (F_{P-}) (left panel) and the contribution of the anomalous scatterer F_A. Only the anomalous components F''_{A+} and F''_{A-} break Friedel`s Law, resulting in different vector sums F_{PA+} and F_{PA-}. As the phase of these is not known we can represent them as circles with a radius of their structure factor magnitudes and their origin being shifted by their corresponding anomalous difference. These two circle intersect twice, giving two possible solutions for the phase. As it is unknown which of the solutions is the correct one, the average of both (F_{SAD}) is used for phasing. (C) 2011 From Biomolecular Crystallography by Bernhard Rupp. Reproduced by permission of Garland Science/Taylor & Francis LLC.

3.2.5.5 Model building and refinement

After initial phase information has been obtained the model of the protein can be built and by this the still erroneous phases can be improved. This is a cyclic process as improved phases lead to a better map which should allow for a more detailed model to be built. During this cyclic process two factors are of importance: the R_{work} and R_{free}. The R_{work} is a measure of the quality of fit of the built model into the electron density. Its value should drop during refinement and it should reach a value of about 15-25%, depending on the quality of the data. Yet, it is possible to fit the electron density with any atoms in a chemically not reasonable way to obtain very low R_{work} values. To avoid this problem, called overfitting, the R_{free} is introduced. It compares the structure factors generated by the model to a small subset of the data (usually 5%) that has not been used for refinement and is thus independent. The R_{free} is

always a few percent higher than the R_{work}, but the gap between them should not exceed 7-8%. The quality of the model is also validated by checking various chemical parameters, such as the dihedral angles in the protein backbone (as presented in the Ramachandran plot), torsion angles and bond length.

3.2.5.6 Refinement and validation

For all proteins refinement was done with Refmac5 as included in the CCP4 suite. Waters were added manually and using the find water option in Coot. TLS groups for the proteins were determined by the *TLS Motion Determination Server* (Painter & Merritt 2006b) and a final round of TLS refinement (Painter & Merritt 2006a) in Refmac5 was carried out. The models were validated using the Ramachandran plot (Ramachandran & Sasisekharan 1968) and the torsion outliers tools in Coot as well as by Molprobity (Chen *et al.* 2010).

3.2.6 Bioinformatics and tools

Plasmids and primers were archived in VectorNTI, which was also used for virtual cloning. The Protparam tool (Gasteiger *et al.* 2003) was used to obtain theoretical data about the protein constructs such as extinction coefficients. Multiple sequence alignments were generated using ClustalW (Larkin *et al.* 2007) as incorporated in the T-coffee server (Notredame 2010) and figures of these were produced with ESpript (Gouet *et al.* 1999). Figures of the protein models were generated with pymol. The PISA server (Krissinel & Henrick 2007) was used to assess the relevance of the protein contacts in the crystal. All protein models were deposited in the Protein databank (PDB) (Berman *et al.* 2002).

4 Results

4A Virulence regulation in pathogenic Yersinia

4A.1 The transcription factor RovA

4A.1.1 Purification of RovA

The plasmid bearing the gene for RovA C81S, C108S (from now on simply termed RovA) was kindly provided by Boris Grujic (HZI, Braunschweig). Several mutants of RovA were generated by site directed mutagenesis and purified as described below for the wild type. The plasmid was transformed into BL21 cells and the protein was produced in M9 medium. After cell lysis RovA was purified from the supernatant via Ni-NTA affinity chromatography. The protein was eluted from the column using the elution buffer containing 250mM imidazole and the elution fractions were analyzed on an SDS gel (Fig. 15).

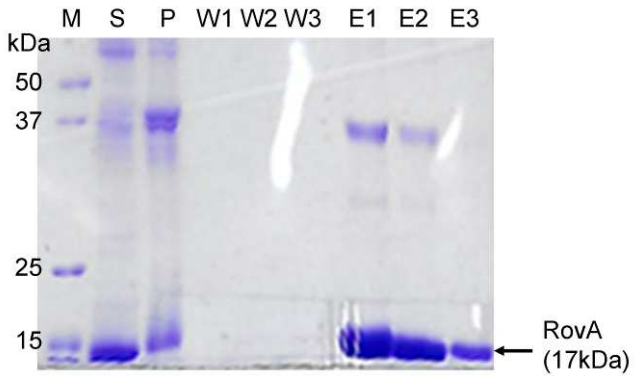


Fig. 15: SDS gel of RovA Ni-NTA chromatography. M: Marker; S: soluble fraction; P: pellet; W1-3: wash; E1-3: elution fractions.

A ThermoFluor assay revealed that RovA is most stable in buffers with high pH, thus a buffer with Tris/HCl pH 9 was used for gel filtration (Fig. 16). The protein eluted from the 16/60 SD200 column as a single peak at about 84ml. This corresponds to a molecular mass of about 40kDa. The calculated molecular mass of RovA is 17kDa and, thus, the calculated oligomeric state is 2.3, indicating a dimer as shown for other MarR type proteins (Hong *et al.* 2005). For further studies the protein was

concentrated to 50mg/ml and frozen in liquid nitrogen. The overall yield of pure RovA was about 15mg/L culture.

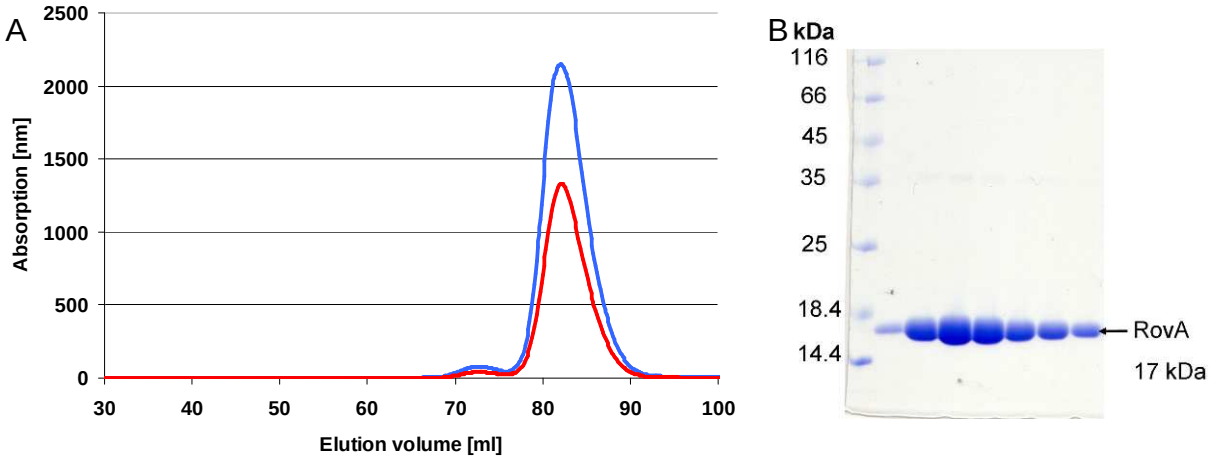


Fig. 16: RovA gel filtration. A: Chromatogram of gel filtration run. Absorption at 280nm is shown in blue, absorption at 260nm in red. B: SDS gel of eluted fractions.

4A.1.2 Thermal stability

RovA has been previously described as a protein thermometer that undergoes a conformational change upon increase in temperature leading to reduced DNA binding affinity and increased proteolytic degradation of RovA (Herbst *et al.* 2009). CD spectrum analysis suggested that RovA is a mostly α-helical protein. The conformational change upon temperature shift leads to a loss of α-helical and a gain of β-sheet content. These findings were investigated in more detail with RovA and several mutants. CD spectra of RovA were recorded in 3.3℃ steps from 10℃ to 37℃ (Fig. 17). The spectra remain relatively similar between 10℃ and 20℃, whereas they change strongly with every interval afterwards. This indicates that the RovA structure remains the same at low temperatures and undergoes a conformational change at higher temperatures. It is however unclear whether all RovA molecules undergo a continuous change which increases upon temperature shift or whether there are only two distinct conformations and the population gradually shifts from one state to the other. Also the molecular mechanism of this conformational change remains elusive. Therefore, we aimed to find mutations in RovA which influence this thermosensing property.

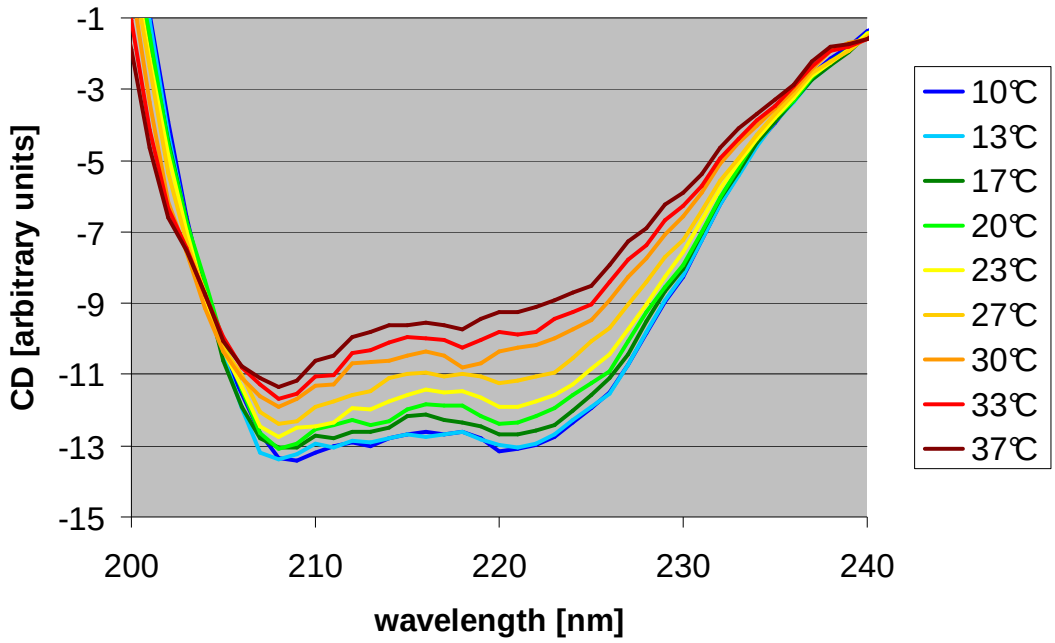
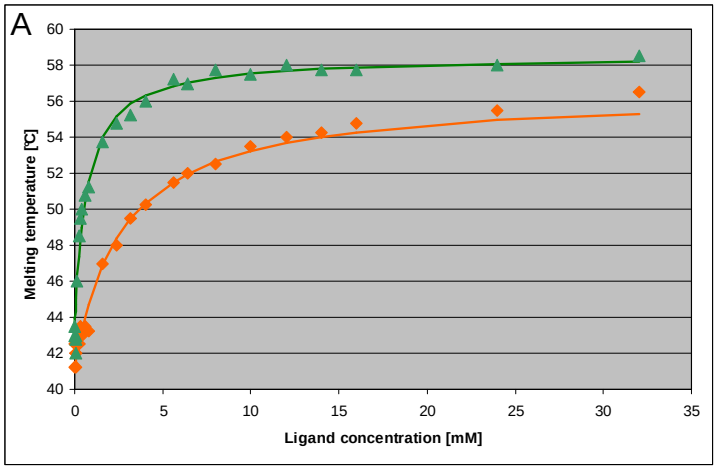


Fig. 17: CD spectrum of RovA at different temperatures.

4A.1.3 Binding of small compounds

Many MarR-type transcription factors have been shown to bind small inducer molecules, most notably salicylate (Kumarevel *et al.* 2009; Providenti & Wyndham 2001; Wilkinson & Grove 2004; Wilkinson & Grove 2006). Yet, for RovA no such inducer molecule has been reported. Thus, salicylate and several similar compounds have been tested by ThermoFluor assay for their capability to bind to RovA (Fig. 18). These experiments show that RovA is able to bind to salicylate with millimolar affinity. The addition of salicylate stabilises RovA by about 15℃. All related compounds such as benzoate and 4-hydroxybenzoate also bind RovA with comparable affinity, while derivatives with amino groups such as anthranilate and 4-aminobenzoate have slightly lower affinities and phenylalanine does not bind at all to RovA in the concentration range tested. Uric acid, which has been identified as a ligand for HucR (Wilkinson & Grove 2004), also does not bind to RovA. Interestingly, RovA showed a higher affinity for cinnamate and 3-chlorocinnamate than salicylate (about 1.8fold and 4.3fold, respectively).

A



B

Compound	K_D [mM]
Salicylate	2,25
Benzoate	2,93
2,3-Dihydroxybenzoate	2,33
Anthranilate	3,92
4-Hydroxybenzoate	2,75
4-Aminobenzoate	5,86
Cinnamate	1,27
3-Chlorcinnamate	0,53
Phenylpyruvate	3,52
Phenylalanine	no binding
Uric Acid	no binding

C

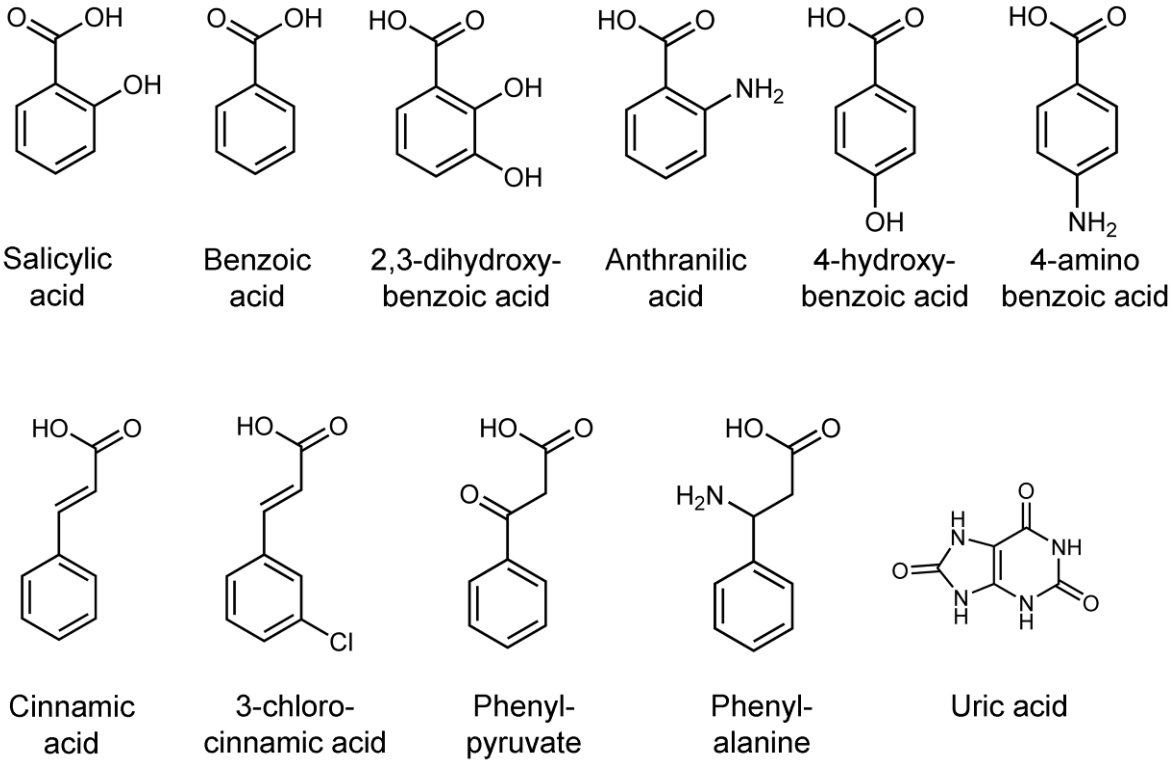


Fig. 18: Affinity of RovA for small molecules. A: Plot of melting temperature against ligand concentration for RovA with salicylate (orange) and 3-chloro-cinnamate (green). B: List of dissociation constants for the tested molecules. C: Structures of tested molecules

To test whether the binding of salicylate to RovA has any impact on the observed conformational change upon temperature change CD measurement with and without 2mM salicylate were performed at 20℃ and 37℃ (Fig . 19). Unfortunately, salicylate absorbs light in the UV range and thus only 2mM salicylate could be used and even then the signal was only usable until about 215nm. Yet, even these truncated data

show that addition of salicylate stabilises the protein and seems to prevent the conformational change.

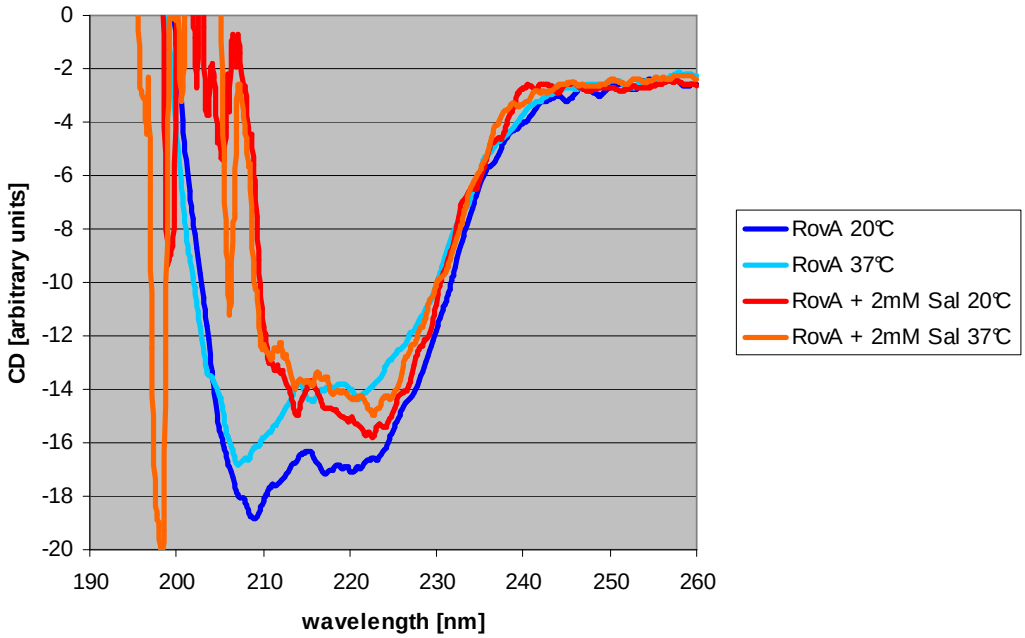


Fig. 19: CD spectra of RovA with and without 2mM salicylate at different temperatures.

The impact of salicylate on DNA binding by RovA has also been tested by Frank Uliczka from the group of Prof. Petra Dersch, HZI. Gel shift assays were performed with varying concentrations of RovA in the presence and absence of 10mM salicylate with a fragment from the *inv* promoter region and, as a control, the *csiD* fragment from *E. coli* (Fig. 20A and B). These gels show that, while RovA alone binds to the *inv* promoter fragment with nanomolar affinity (a K_D of 32nM has been reported (Herbst *et al.* 2009)), DNA binding seems to be severely impaired in the presence of salicylate. About twice as much RovA was needed compared to the gel without salicylate to observe a shift of the DNA. These results indicate that RovA is able to bind salicylate which stabilises RovA resulting in a reduced conformational flexibility and an increased melting point as well as decreases its DNA binding affinity.

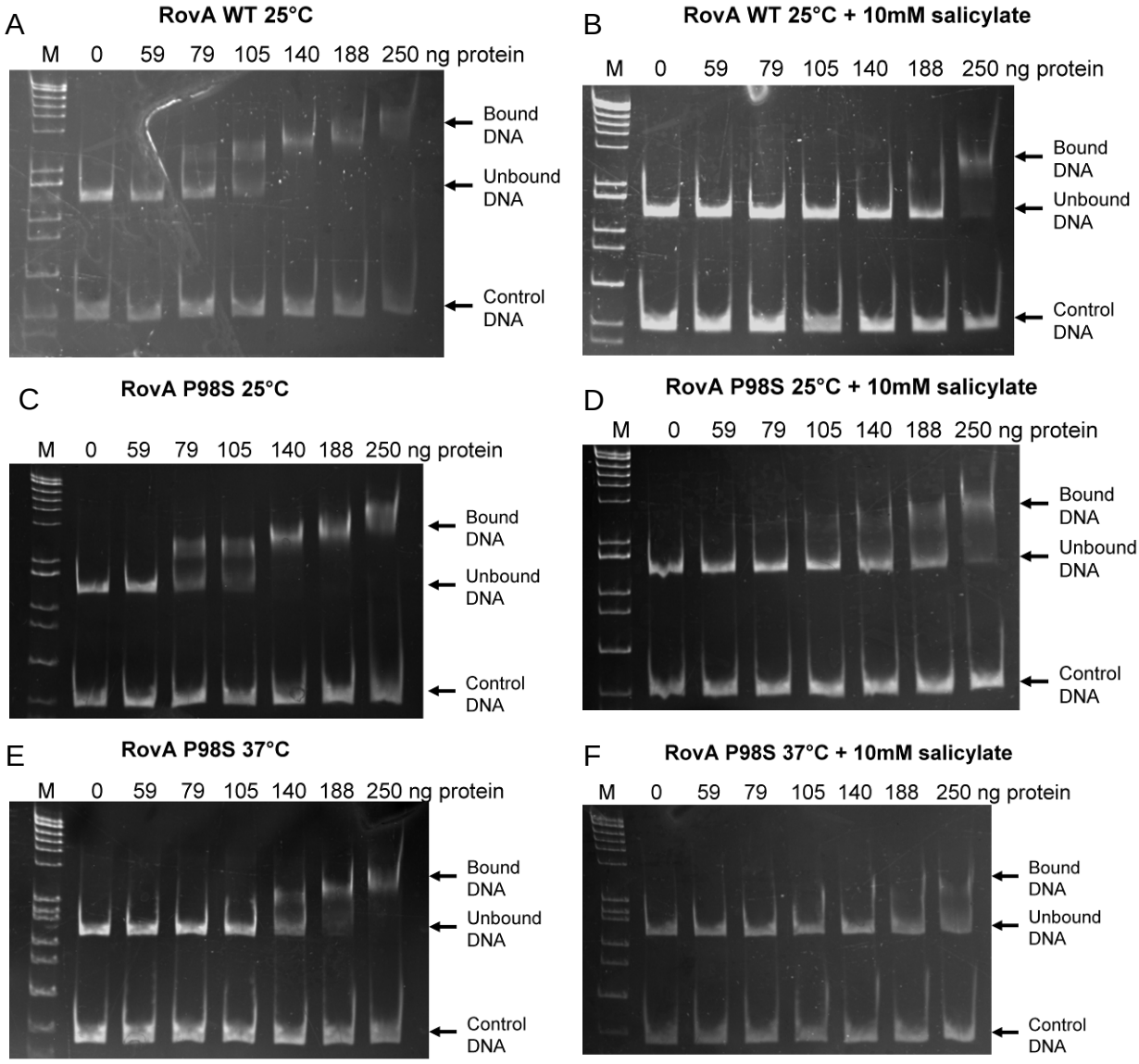


Fig. 20: Gel shift assays of RovA and RovA P98S at 25℃ and 37℃ and in the presence and absence of 10mM salicylate.

In order to determine the effect of salicylate-induced stabilisation *in vivo*, *Y. pseudotuberculosis* cells were grown in the presence of 1mM salicylate and the degradation of RovA with time was monitored by western blots. Preliminary results by Chriselle Mendonca as part of her PhD thesis indicate that salicylate is able to stabilize RovA *in vivo* (Fig. 21). Significantly more RovA was visible directly after blocking synthesis as well as later. Taken together, these results suggest that a small molecule such as salicylate might be involved in the regulation of RovA by stabilizing it.

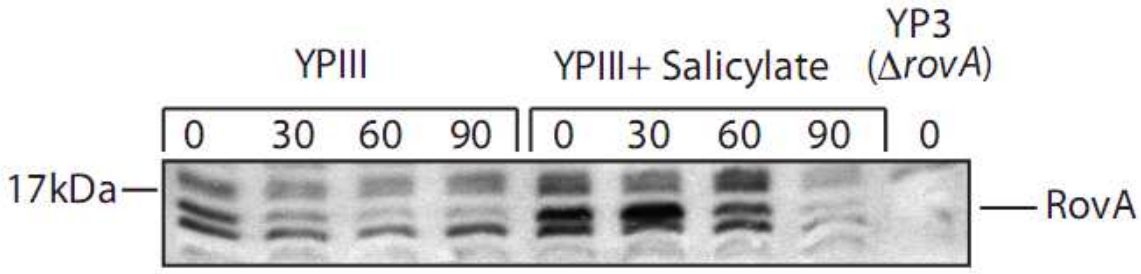


Fig. 21: Stability assay of RovA in *Y. pseudotuberculosis* cells. YPIII cells were grown in the presence and absence of 1mM salicylate. Protein synthesis was stopped and the amount of RovA was monitored by a western blot. In the presence of salicylate significantly higher amounts of RovA can be seen in comparison to the absence of salicylate.

4A.1.4 RovA P98S is more stable than the wild type *in vivo*

A mutant of RovA (P98S) has been found by Frank Ulizcka in isolates of *Y. enterocolitica* serotype O:3 which showed an increased stability at 37℃ *in vivo* compared to the wild type. Samples of *Y. pseudotuberculosis* YP3 and YP67 (a Δlon mutant strain) complemented with a plasmid for the production of RovA and RovA P98S at 25℃ and 37℃ were analyzed on a western bl ot (Fig. 22). These blots indicate that RovA P98S is much more stable at 37℃ than the RovA WT. To test whether this effect is due to a difference in DNA binding, inducer binding, thermal or proteolytic stability gel shift assays with RovA P98S were performed at 25℃ and 37℃ and in the presence or absence of salicylate (Fig. 20C-F). These gels show that RovA P98S binds the inv promoter with similar affinity as the wild type and that the DNA binding is hampered by the addition of salicylate. Additionally, DNA binding affinity is also affected by increased temperature like the wild type.

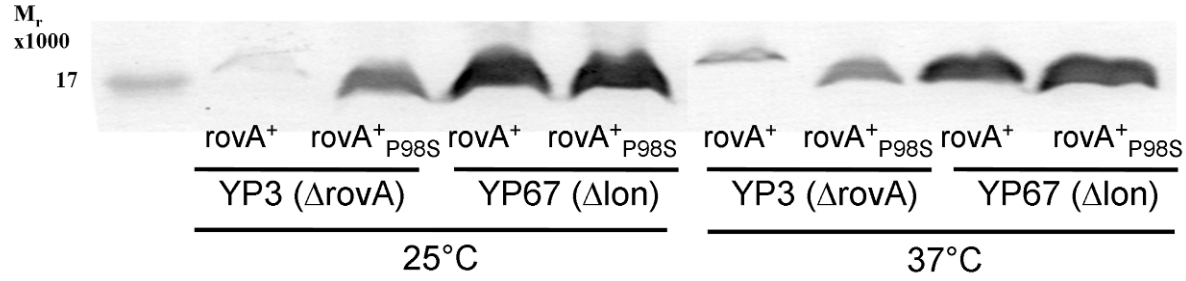


Fig. 22: Complementation assay of RovA WT and RovA P98S in YP3 and YP67 (Δlon) background. Samples from these strains producing the RovA variants at different temperatures were analysed on a western blot.

We further examined this mutant by thermofluor assays and CD spectroscopy. With the ThermoFluor experiment no difference in the melting temperature of PovA P98S compared to the wild type could be detected (35°C for both; data not shown). The CD spectra showed that RovA P98S exhibits the same changes upon temperature shift as the wild type (Fig. 23). All these results together suggest that the mutation P98S in RovA does not influence the ability to bind DNA or salicylate nor influences the thermal stability of the protein. Therefore, we assume that this mutation has an influence on the proteolytic stability *in vivo*.

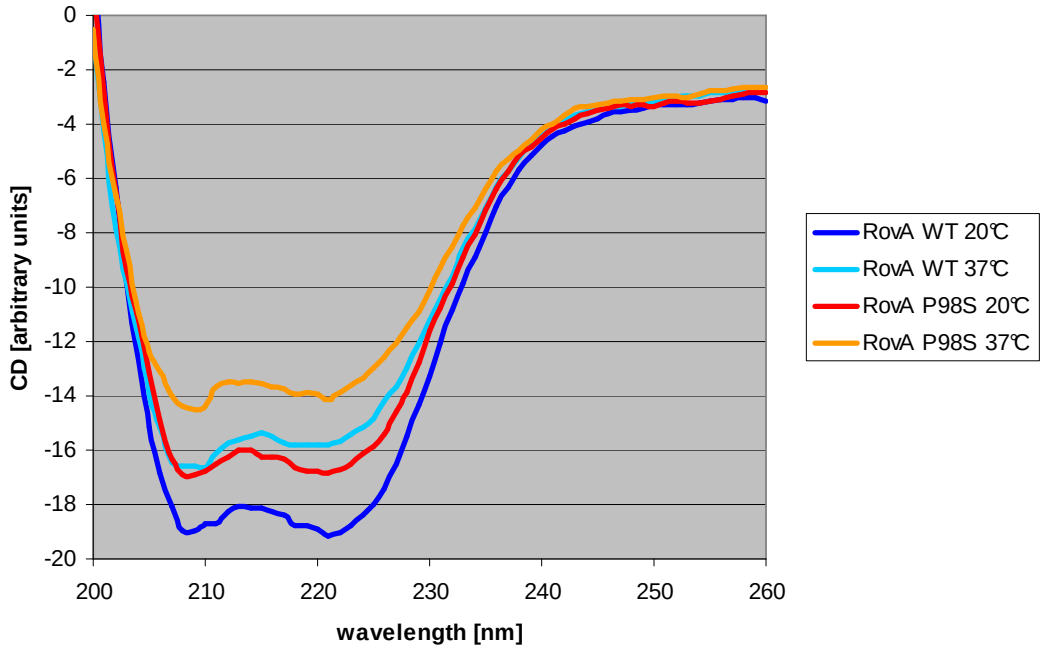


Fig. 23: CD spectra of RovA WT and RovA P98S.

4A.1.5 RovA A10F, a stable mutant *in vitro*

In order to interfere with the observed ligand binding, mutants of RovA (G6F and A10F) were produced in which small residues lining the proposed ligand binding pocket were replaced by the large amino acid phenylalanine. The stability and salicylate binding of these mutants were measured by ThermoFluor measurement (Fig. 24A). RovA G6F showed a similar melting temperature in the absence of salicylate as the wild type and was stabilised by the addition of salicylate like the wild type (K_d = 2mM and 2.4 mM for RovA G6F and the wild type, respectively). On the other hand, RovA A10F showed a significantly increased melting temperature in the absence of salicylate (T_M = 53°C). RovA A10F is also stabilised by salicylate, albeit to

a lesser extend than the wild type. The calculated dissociation constant for RovA A10F is 8mM, about 3 times higher than for the wild type.

The mutants were also tested for their conformational flexibility by CD measurements at different temperatures. Again, RovA G6F shows spectra which are very similar to the wild type (data not shown), whereas RovA A10F showed a considerably reduced difference between the spectra recorded at 20℃ and 37℃ than the wild type (Fig. 24B). These results suggest that the mutant RovA A10F is much more stable *in vitro* than the wild type and less able to bind salicylate.

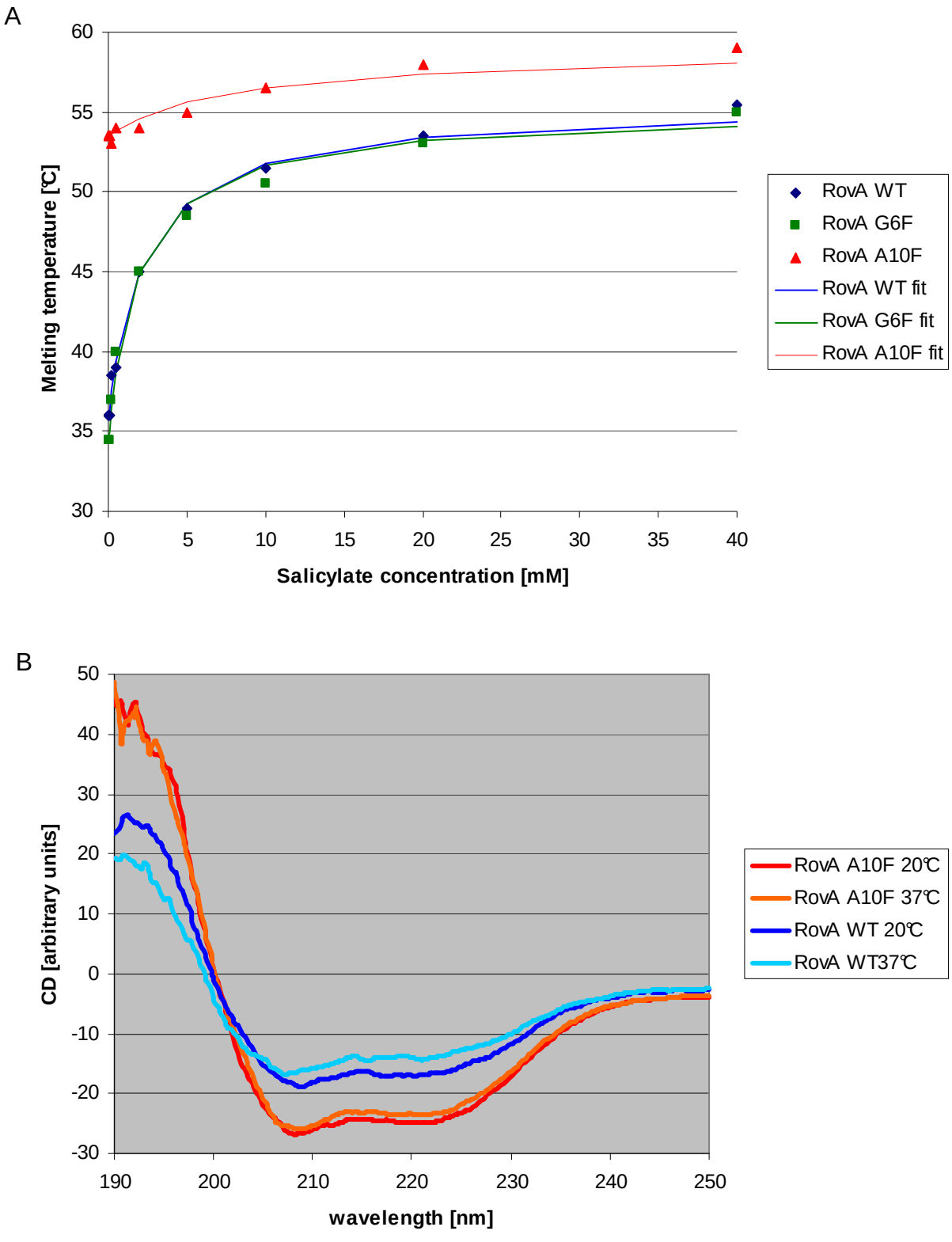


Fig. 24: Stability and salicylate binding of RovA wild type and G6F and A10F mutants. A: The melting temperature was measured by a ThermoFluor experiment in the presence of increasing concentrations of salicylate and plotted against the salicylate concentration. B: CD spectra of RovA wild type and RovA A10F recorded at 20℃ and 37℃.

4A.1.6 Crystallisation of RovA

Crystallisation of RovA-apo was done by Boris Grujic as part of his PhD thesis (personal communication). RovA readily crystallizes in a variety of PEG containing conditions such as 0.1M MES pH 6.5, 25% PEG2000MME (Fig. 25A). Yet, even though the crystals can grow to a large size (up to 500µm) they do not diffract very well. Microseeding resulted in even larger crystals which were unfortunately not single crystals (Fig. 25B). For data collection the crystals were transferred into a cryo buffer containing 0.1M MES pH 6.5, 30% PEG2000MME and 10% glycerol and then frozen in liquid nitrogen.

A B

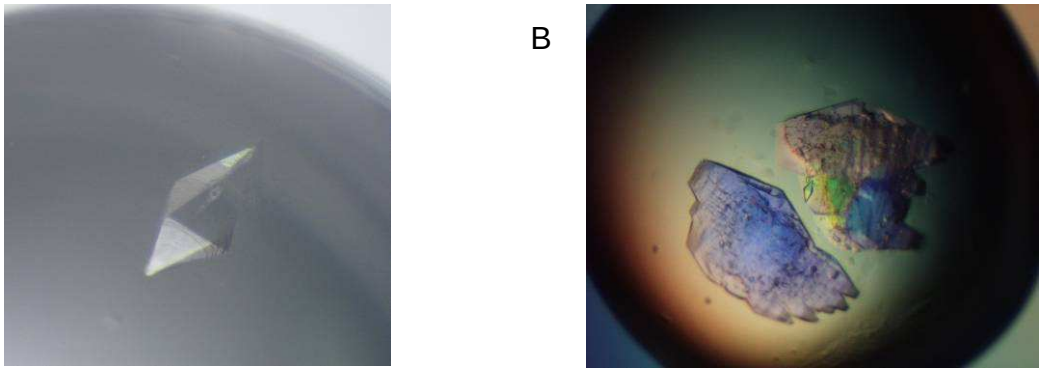
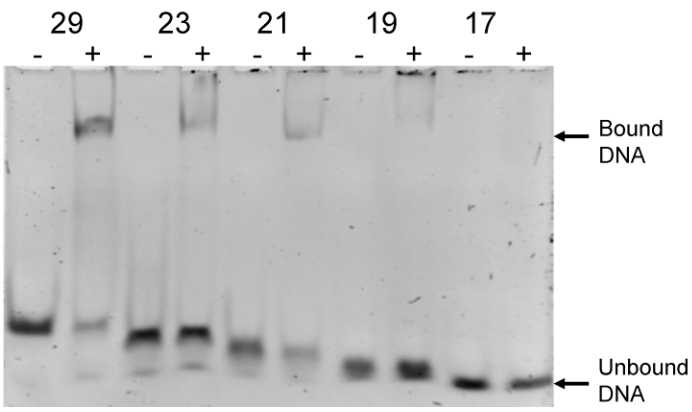


Fig. 25: Crystals of RovA apo. A: Non-seeded crystal B: Seeded crystals

4A.1.7 Determination of DNA fragment for cocrystallization

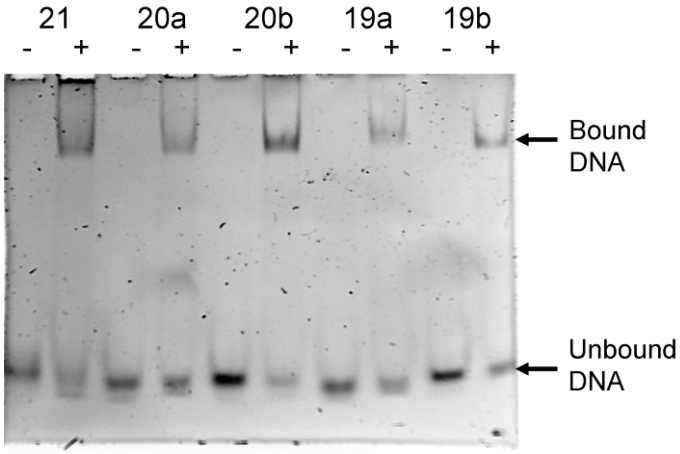
The success of protein/DNA cocrystallization is strongly dependent on the length and overhang of the DNA fragment (Tan *et al.* 2000) as the DNA fragments often assemble in a head to tail fashion to form a pseudo-continuous helix throughout the crystal. Therefore, the minimal DNA fragment still able to bind RovA efficiently was determined by gel shift assays. The smallest RovA binding fragment that had been determined previously from the *inv* promoter was the 29bp sequence 5`-taattgatattatttatatgataatagtt -3` (called 29; of the blunt end double stranded DNA, only one strand is shown here). From this smaller fragments were generated by deleting one base pair at each side of the fragment (leading to 27-11) and tested for RovA binding. The gels showed that the DNA binding capacity of the fragments decreased with the size of the fragment and that 21 was the shortest fragment that still bound RovA strongly (Fig. 26).



29: 5`- taattgatattatttatatgataatagtt - 3`
27: 5`- aattgatattatttatatgataatagt - 3`
25: 5`- attgatattatttatatgataatag - 3`
23: 5`- ttgatattatttatatgataata - 3`
21: 5`- tgatattatttatatgataat - 3`
19: 5`- gatattatttatatgataa - 3`
17: 5`- atattatttatatgata - 3`
15: 5`- tattatttatatgat - 3`
13: 5`- attatttatatga - 3`
11: 5`- ttatttatatg - 3`

Fig. 26: Gel shift assay of RovA with DNA fragments of varying size. +/- indicate the presence or absence of RovA. On the right one strand of the used fragments is shown, the complementary strand is left out for clarity.

To test this fragment in more detail, new fragments were tested in which one basepair at only one side was deleted (Fig. 27). This experiment showed a preference of RovA for 3`-end truncations and thus 19b with the sequence 5`- tgatattatttatatgata - 3` was used as a minimal fragment.



21: 5`- tgatattatttatatgataat - 3`
20a: 5`- gatattatttatatgataat - 3`
20b: 5`- tgatattatttatatgataa - 3`
19a: 5`- atattatttatatgataat - 3`
19b: 5`- tgatattatttatatgata - 3`

Fig. 27: Gel shift assay of RovA with DNA fragments of varying size. +/- indicate the presence or absence of RovA. On the right one strand of the used fragments is shown, the complementary strand is left out for clarity.

4A.1.8 Crystallisation of RovA in complex with DNA

From the determined minimal fragment 19b 5`- tgatattatttatatgata - 3` from the inv1 site six different double stranded DNA fragments with different lengths and overhangs were devised. Fragments 20 and 21 were blunt end fragments, while 20A and 25A have an overhang for Watson-Crick base pairing. 20A allows the packing of

linear DNA, whereas 25A allows for bent DNA (Tan *et al.* 2000). Fragments 19C and 19G were designed to allow Hoogstein base pairing. Yet, of these only 20A yielded crystals (see below). To obtain further insights into the DNA binding of RovA other fragments were designed in the fashion of 20A from two binding sites in the rovA promoter and from the second RovA binding site on the inv promoter (Table 8). Of these only Inv1 Art and rovA1A yielded crystals.

Table 8: DNA fragments used for cocrystallization with RovA

Name	Sequence	Overhang
Inv1 20:	FW: 5` - tgatattatttatatgataa - 3`	Blunt end
	RV: 3` - actataataaatatactatt - 5`	
Inv1 21:	FW: 5` - ttgatattatttatatgataa - 3`	Blunt end
	RV: 3` - aactataataaatatactatt - 5`	
Inv1 20A:	FW: 5` - atgatattatttatatgataa - 3`	
	RV: 3` - actataataaatatactattt - 5`	5`A
Inv1 25A:	FW: 5` - aaattgatattatttatatgataata - 3`	
	RV: 3` - ttaactataataaatatactattatt - 5`	5`A
Inv1 19C:	FW: 5` - cgatattatttatatgatc - 3`	
	RV: 3` - ctataataaatatactagc - 5`	5`C
Inv1 19G:	FW: 5` - ggatattatttatatgatc - 3`	
	RV: 3` - ctataataaatatactagg - 5`	5`G
Inv1 Art:	FW: 5` - atatgataaatgatattattt - 3`	
	RV: 3` -atatactatttactataataa- 5`	3`A
Inv2a:	FW: 5` - aatagactgttttattatata - 3`	
	RV: 3` - tatctgacaaaataatatatt- 5`	5`A
Inv2b:	FW: 5` - aataaaacagtctataccata - 3`	
	RV: 3` - tattttgtcagatatggtatt - 5`	5`A
Rova1a:	FW: 5` - aattatattatttgaattaat - 3`	
	RV: 3` - taatataataaacttaattat- 5`	5`A
Rova1b:	FW: 5` - tttacatccattaattatata - 3`	
	RV: 3` - aatgtaggtaattaatatatt - 5`	5`A
Rova2a:	FW: 5` - tatgctagcacgctaattaaa - 3`	
	RV: 3` - tacgatcgtgcgattaatttt- 5`	5`A

RovA was crystallized in complex with DNA at a concentration of 20mg/ml protein and 8mg/ml DNA. Initial crystals appeared under the condition 0.05 M MES pH 5.6, 0.01 M $MgCl_2$, 2 M Li_2SO_4 (Fig. 28A). Optimisation of the crystallisation condition resulted in large, hexagonal crystals which strongly polarize the light, an indication for DNA, under the condition 0.05 M MES pH 5.6, 0.01 M $MgCl_2$, 2.3 M Li_2SO_4 (Fig. 28B). The crystals appeared after 1 week and grew to full size within 3 weeks. The crystals were transferred into a cryo buffer containing 0.05 M MES pH 5.6, 0.01 M $MgCl_2$, 2.5 M Li_2SO_4 and 20% glycerol and then frozen in liquid nitrogen.

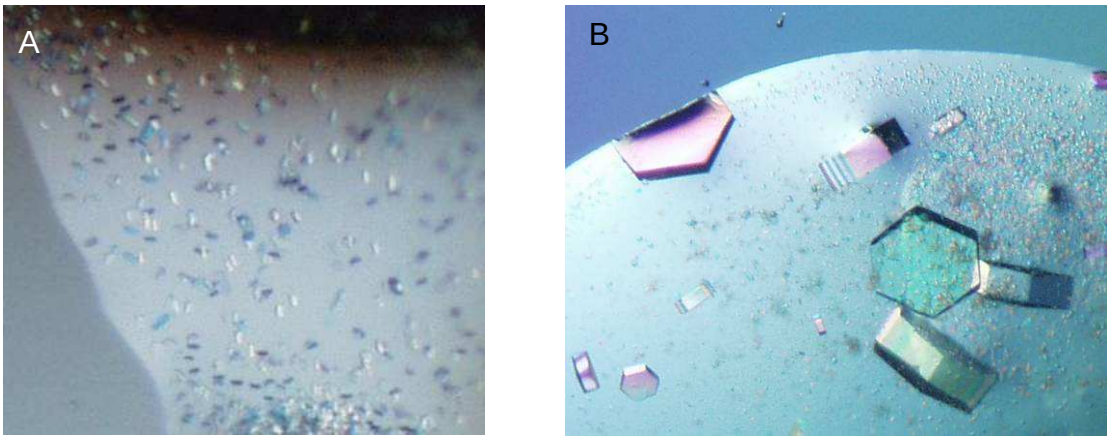


Fig. 28: Crystals of RovA in complex with DNA. A: Initial crystals, B: Optimised crystals.

4A.1.9 Crystallization of RovA with salicylate

RovA was crystallized together with 50mM salicylate at a protein concentration of 20mg/ml. Under several conditions, such as 0.1M MES pH 5, 2.4M ammonium sulfate, bundles of needles were observed (Fig. 29A). These crystals could be reproduced in 24-well hanging drop plates, yet, even with seeding it was not possible to obtain single crystals. Only by crystallization at 4℃, incubating for two days and then streak seeding under the condition 0.1M MES pH 5.6, 2 M ammonium sulfate, some single crystals could be obtained. Before freezing the crystals were transferred into a solution containing 0.1M MES pH 5.6, 2.5 M ammonium sulfate and 15% glycerol.

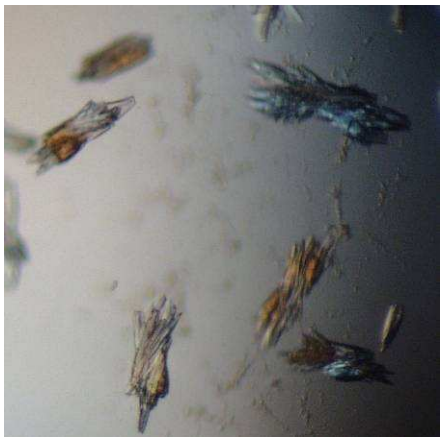
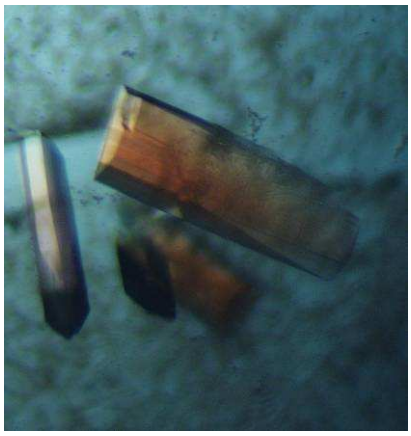
A  B 

Fig. 29: Crystals of RovA with salicylate. A: Initial crystals; B: Optimised crystals

4A.1.10 Data collection

The RovA-apo crystals belong to the space group $P2_12_12_1$ and a dataset up to 2.4 Å was collected at the ESRF ID14h2. The crystals of RovA in complex with DNA belong to the space group P3 and diffracted up to 1.85 Å, whereas the crystals of RovA in complex with salicylate belong to the space group C2 and diffracted up to 2.4 Å The datasets were indexed with XDS (Kabsch, 2010) and scaled with Scala (Evans, 2006) as included in the CCP4 suite (CCP4, 1994). The Mathews coefficient for the RovA-apo data (Matthews, 1968) was determined to V_M = 2.25 $Å^3$ Da^{-1}, indicating three RovA dimers in the unit cell. For the RovA-DNA crystals the Mathews coefficient (Matthews, 1968) was determined to V_M = 3.33 $Å^3$ Da^{-1}, indicating a RovA dimer bound to one double-stranded DNA oligo in the unit cell, and V_M = 1.99 $Å^3$ Da^{-1} for the salicylate containing crystals, indicating a RovA dimer in the unit cell. Molecular replacement was performed using MrBump (Keegan & Winn, 2007) employing Molrep (Murshudov et al., 1997) using SlyA from *Salmonella typhimurium* (accession code 3deu) as a model. However, molecular replacement for RovA-apo was only succesful using the structure of RovA in compex with the DNA. The protein models were built in Coot (Emsley et al., 2010) and refined using Refmac (Murshudov et al., 1997). Waters were added by the water finding tool in Coot as well as manually. In the end, a final refinement step using TLS refinement (Painter & Merritt, 2006a) with TLS groups as defined by the TLS Motion Determination Server (Painter & Merritt, 2006b) was carried out. In case of the DNA-complex crystals unambiguous density for the DNA was found. The DNA was modelled into the density and after refinement the Rwork and Rfree finally converged to 21% and 24% respectively. R_{work} and R_{free} for the RovA-apo structure converged to 22.8 and

28.4%, respectively. For the salicylate containing crystals, regions of electron density were found in four sites in the protein, which could be well described by salicylate. As the work on the RovA salicylate data could not be finished, the Rwork and Rfree were relatively high with 31.4% and 35.4%, respectively. This might be due to problems with crystal twinning.

Table 9 and Table 10 summarize the data collection and refinement statistics. The final models were validated using MolProbity (Davis et al., 2007). For all models, all the residues fall into the allowed region of the Ramachandran plot (Ramachandran & Sasisekharan, 1968).

Table 9: RovA/ DNA data collection and refinement statistics

	RovA + inv1	RovA + rova1
X-ray source	ESRF ID14eh4	ESRF ID14eh4
Space group	P3	P3
Unit cell dimensions *a*, *b*, *c* (Å)	88.7, 88.7, 67.5	88. 8, 88.8, 67.7
Wavelength (Å)	0.95	0.95
Resolution (Å)*	76.8 – 1.85	37.1 – 2.1
Measured	127340 (18753)	209176 (13305)
Unique reflections*	49796 (7414)	39678 (5287)
I/σ_I*	13.9 (2.5)	15.3 (2.0)
Completeness (%)*	98.2 (99.9)	97.6 (89)
Redundancy*	2.6 (2.5)	5.3 (2.5)
R_{merge} (%)*	3.4 (48.7)	5.8 (45.7)
Wilson B factor (Å^2)	33.74	38.8
Solvent content (%)	69%	69%
R_{work} (%)	19.7	21.3
R_{free} (%)	23.1	25.2
Protein atoms	2196	2162
DNA atoms	855	855
Solvent atoms	187	76
r.m.s.d. from ideal geometry:		
-Bond lengths (Å)	0.011	0.022
-Bond angles (°)	1.44	2.39
Average B-factor		
-Protein (chain A)	17.1	23.5
-Protein (chain B)	17.8	21.9
-DNA (chain C)	33.1	43.6
-DNA (chain D)	32.2	45.8
-Solvent	20.6	19.6
Ramachandran plot (%):		
-Favoured region	98.6	95
-Allowed region	1.4	5

* Values in parentheses refer to statistics in the highest resolution shell

4A.1.11 RovA/DNA crystal packing

RovA was successfully crystallized in complex with a fragment of DNA from the *inv* promoter region (inv1) and another fragment from the *rovA* promoter (rovA1A). The structure was solved by molecular replacement using the homologue SlyA (accession code 3DEU), which shares a 75% sequence identity, as a model. The resulting map showed clear density for DNA as well as the loop between residue 80 to 89, both of which were absent in the replacement model. In the space group P3 there is one RovA dimer bound to one DNA fragment. In the crystal RovA is arranged layers where one dimer interacts with 4 other RovA dimers (Fig. 30). The RovA layers do not touch each other and crystal contacts between them are only mediated by the bound DNA which lies perpendicular to these planes. These DNA fragments bind to one another with their one basepair overhang in a head to tail fashion to yield a pseudo continuous DNA strand which spans the whole crystal. This crystal packing explains the high solvent content of 69%.

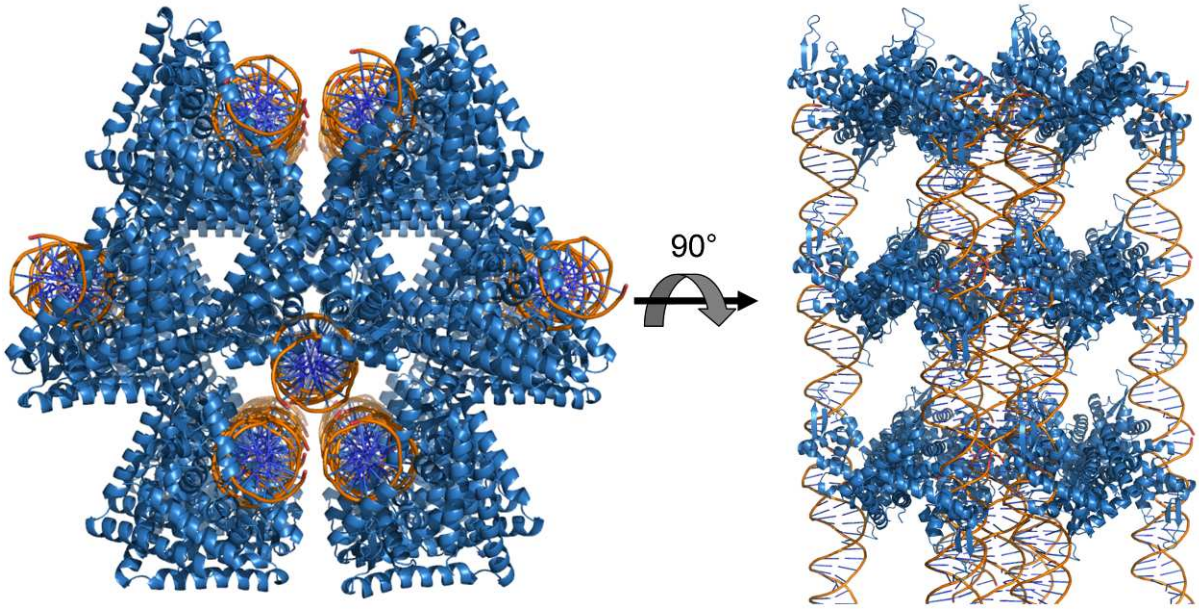


Fig. 30: RovA/DNA crystal packing as viewed from above and from the side. RovA is organized in layers and interacts with four other dimers, whereas the layers are only connected by the DNA.

4A.1.12 RovA structure in complex with DNA

The structure of RovA overall resembles that of other MarR type proteins. It consists of six α-helices and two β-sheets with the topology: α1 (residues 5–23), α2 (residues 30–42), α3 (residues 48–55), α4 (residues 60–72), β1 (residues 76–80), β2 (residues

88–92), α5 (residues 98–117) and α6 (residues 120–143) (Fig. 31A and B). From residue 94 to residue 97 as well as from residue 24 to 27 there are two one-turn 3_{10}-helices. In several other MarR type proteins such as OhrR and MarR there is another β-sheet between α_2 and α_3, but in that region in RovA there are two prolines (P44 and P45) that probably prevent the formation of a β-sheet. The two monomers are tightly intertwined with the three long helices α1, α5 and α6, which form the dimer interface with a covered area of 2252.6 $Å^2$ and a ΔG of -34.4 kcal/mol as calculated by the PISA server (Krissinel & Henrick 2007). The DNA binding domain is made up by the central part of RovA, namely helices α2-4, which form a helix-turn-helix (HTH) motif, and β1 and β2 which form a so called wing, to build an overall winged helix-turn-helix (wHTH) domain.

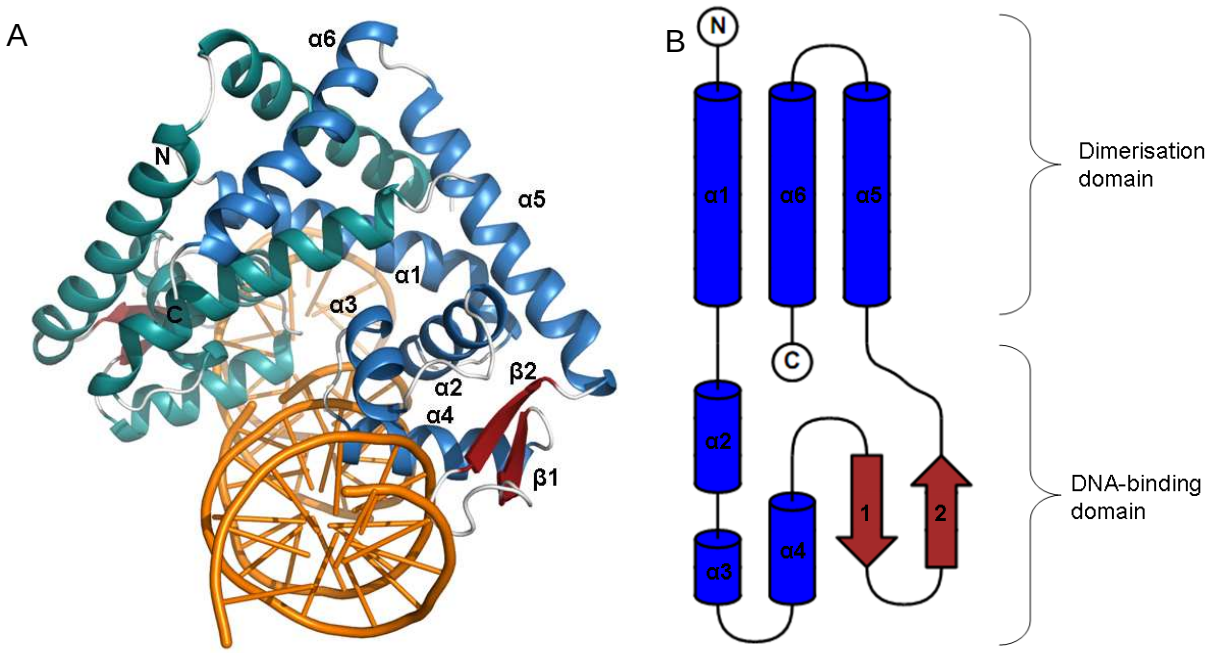


Fig. 31: Structure of RovA. A: Cartoon representation of RovA in complex with DNA. One monomer is shown in blue, the other in blue-green, β-strands are shown in red and DNA in orange. B: Topology plot of RovA, showing domain organisation generated by Topdraw.

4A.1.13 RovA-DNA interactions

RovA was cocrystallized with DNA fragments from the inv (inv1) and rovA (rovA1A) promoter regions. Surprisingly, in the structure of RovA in complex with inv1 RovA does not bind in the middle of the DNA fragment, but rather on the edge of one fragment and also to the next fragment in the pseudo-continuous DNA helix (Fig. 32A). Upon closer inspection of the new binding site sequence, it became evident

that, because of the pseudo-continuous DNA helix in the crystal, a binding site was generated, which is very similar to the original sequence (Fig. 32B). This new binding sequence (inv1new) contains about one half of one DNA fragment and about half of the neighboring fragment. Therefore, it seems that RovA has shifted its binding register by moving along the DNA during crystal assembly probably due to a higher affinity to the new sequence. That this preferential binding to the end of the DNA is not a crystal artifact is proven by the fact that RovA binds to the rovA1A DNA, which has been derived from the *rovA* promoter region, in the middle as expected (Fig. 33A). Overall both structures are very similar with an rmsd between common C_α atoms of the structures of 0.41Å.

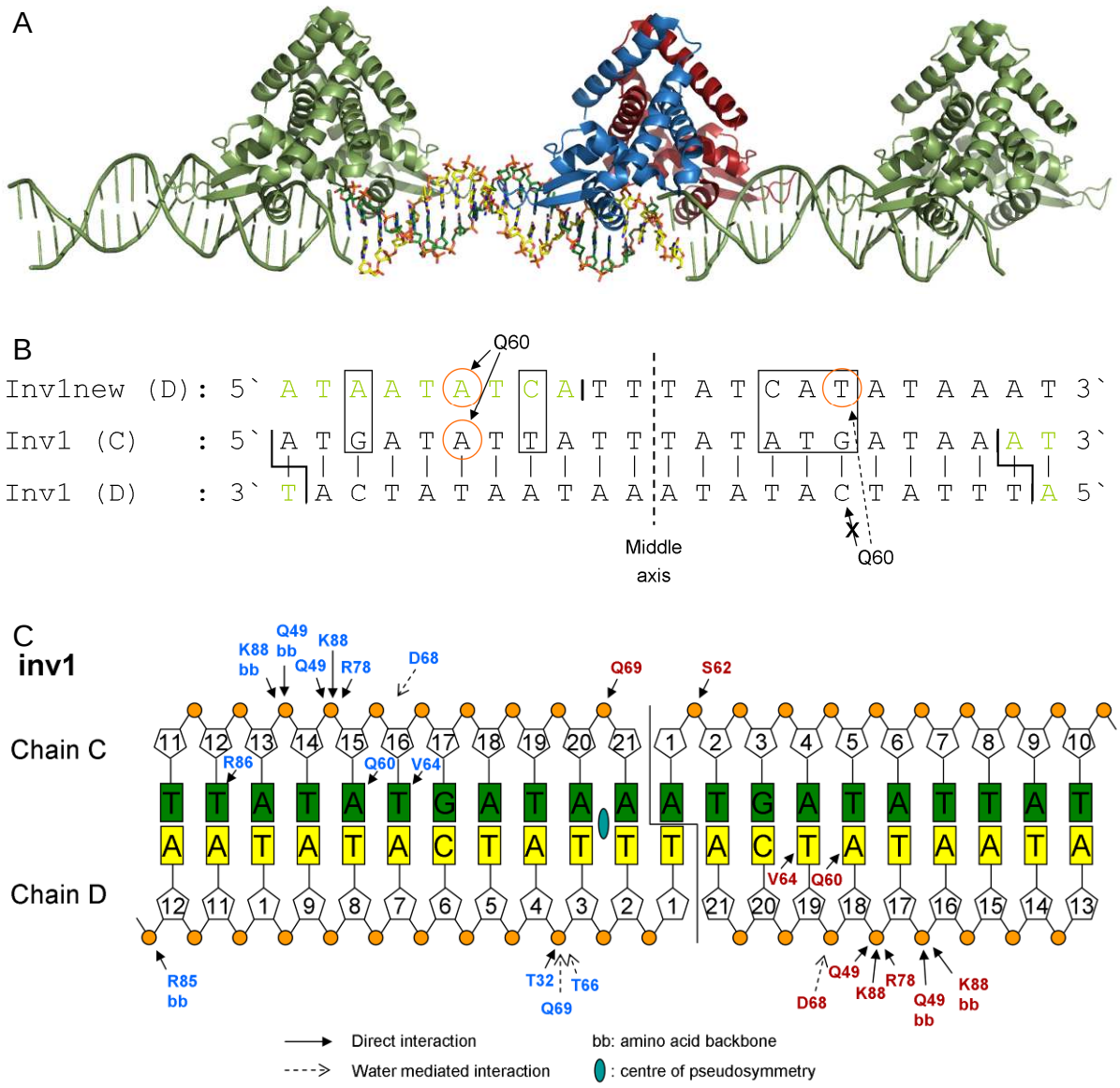


Fig. 32: Binding of inv1 by RovA. A: Cartoon representation of RovA in complex with inv1 DNA. One RovA dimer is shown in blue (chain A) and red (chain B), the DNA in green (chain C) and yellow (chain D) sticks. Two symmetry mates are shown in green. B: The sequence of the DNA fragments inv1 (chain C and D) are shown in black, the sequence of the next symmetry mates in the crystal are shown in green. Chain C was aligned with chain D (inv1new) to show the high similarity of the new binding site with the proposed binding site. Unmatched bases are marked by square boxes, orange circles mark the basepairs recognized by Q60. The dashes line is the middle of the pseudopalindromic sequence. C: Schematic drawing of the RovA inv1-DNA interactions. Direct interactions are symbolized by full arrows, water mediated interactions by dashed arrows. Bases are represented by rectangles, desoxyribose by pentagons, phosphates by circles. The colors of the RovA residues correspond to the chains as in A (blue: chain A, red: chain B).

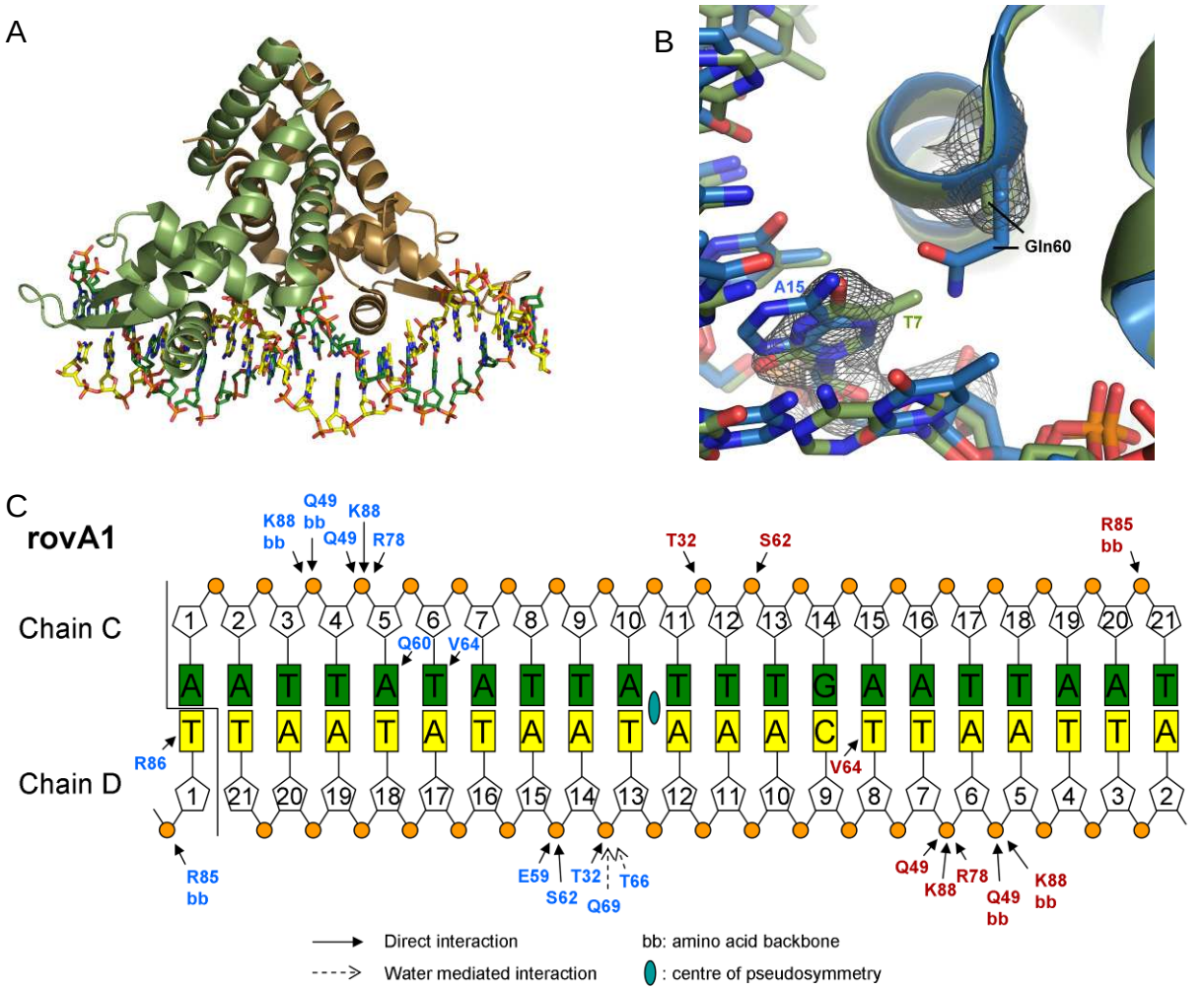


Fig. 33: Binding of rovA1A by RovA. A: Cartoon representation of RovA in complex with inv1 DNA. One RovA dimer is shown in green (chain A) and brown (chain B), the DNA in green (chain C) and yellow (chain D) sticks. B: Structural alignment of RovA bound to inv1 (blue) and rovA1A (green). The role of Q60 is shown as it only binds to adenine. If there is a thymidine, as in rovA1A, the sidechain is disordered. The electron of the rovA1A structure is shown as grey mesh. C: Schematic drawing of the RovA rovA1A-DNA interactions. Direct interactions are symbolized by full arrows, water mediated interactions by dashed arrows. Bases are represented by rectangles, desoxyribose by pentagons, phosphates by circles. The colors of the RovA residues correspond to the chains (blue: chain A, red: chain B).

RovA makes extensive interactions with the DNA in both structures (Fig. 32C and Fig. 33C). Helix α4 of the HTH motif is deeply inserted into the major groove of the DNA and makes most of the contacts to the DNA, but also helices α2 and α3 provide some contacts. The wing protrudes from the protein and reaches into the next minor groove thereby interacting with the DNA at several amino acid sites.

Most of these interactions, however, are not base specific. Instead, mainly the phosphate backbone of the DNA is contacted by hydrogen bonding with amino acids such as K88, Q49 and R78 (Fig. 34A) or backbone nitrogens (Fig. 34B).

Only very few interactions are visible between RovA and the bases of the DNA. The most notable interaction is between Q60 and adenine residues at position 6 from the pseudosymmetric centre (Fig. 34D and Fig. 32C). This interaction might be the reason why RovA did not bind to the inv1 fragment in the middle as expected, as this arrangement would have allowed the binding of Q60 from only one monomer to the DNA, whilst the other would not have an adenine residue to bind to. The way inv1 was found to be bound in the crystal allows both Q60 residues to bind to an adenine residue. Interestingly, in the structure with rovA1A only one Q60 is able to bind the DNA, while the other cannot because there is a thymidine instead of an adenine. For this glutamine residue no density is visible and it is thus probably disordered (Fig. 33B).

Another important residue is probably V64. This makes hydrophobic interactions with a thymidine residue at position 5 from the centre of pseudosymmetry in both monomers of both structures (Fig. 32C and Fig. 33C). Additionally, R86 contacts a thymidine residue while being held in place by D64 by hydrogen bonds (Fig. 34C). However, this interaction has been found only for one monomer of the RovA/inv1 structure, while it was not found in the RovA/rovA1A structure even though there is a thymidine residue in the same position. This may be due to slight differences in the DNA due to the different gap positions.

4A.1.14 A potential ligand cavity contains unexplained density

A small cavity is present in the dimerisation domain of both monomers of RovA. This cavity is mostly lined with hydrophobic residues such as W16, W34, Y38 and I107 (Fig. 34F). Inside this cavity, strong density was visible that could not be atributed to any of the substances from the crystallization buffer. It is therefore possible that a small molecule was bound to the protein from the expression host. The cavity has also been described for other MarR-type proteins suggesting a potential role as an inducer binding pocket (Wilkinson & Grove 2006).

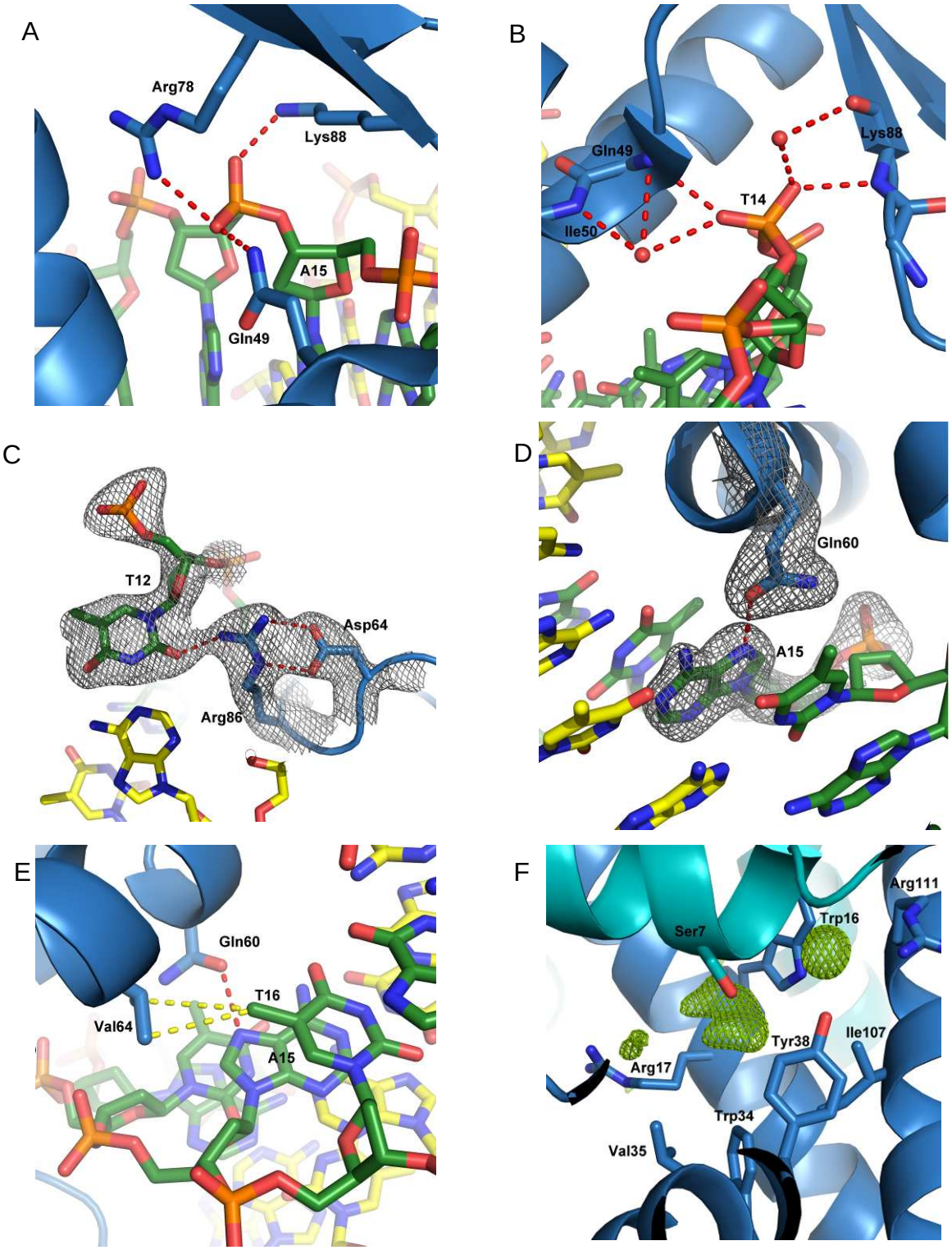


Fig. 34: Critical RovA inv1-DNA interactions. RovA is shown in blue as cartoon and sticks. The DNA is shown as sticks; chain C is green, chain D is yellow. Density is shown at 1 sigma as grey mesh. A: Example of phosphate backbone binding by N49, K88 and R78 of RovA. B: The DNA phosphate backbone is also contacted by RovA backbone nitrogens and by water mediated interactions, here via K88, N49 and I50. C: Basespecific interaction between thymidin 12 and R86 which is stabilized by D64. D: Basespecific interaction between adenine 15 and Q60. E: Hydrophobic interaction between V64 and thymidin 16 (yellow dots). F: Unexplained density within RovA, the F_O-F_C map is shown as green mesh at 4 sigma.

4A.1.15 The RovA-apo structure shows conformational flexibility

By molecular replacement using the RovA/DNA structure five RovA monomers (chains A-E) were found in the asymmetric units. During model building, unassigned density for three long helices was discovered. These could be modelled and monomer A was used for replacing them by superposition. However, the electron density of this chain (F) remained weak in the parts now involved in dimer formation and therefore could be built partially. This is probably due to missing interactions in the crystal lattice. The wing of the DBD is disordered in all monomers of the RovA-apo structure. It is probably very flexible and assumes a stable conformation only in the complex with the DNA.

In total, six monomers were built, forming three dimers. All monomers share the same ovarall fold with rmsd values for common C_α atoms relative to chain A of 1.35 Å/ 0.4 Å/ 1.1 Å/ 0.6 Å/ 0.6 Å for chains B/C/D/E and F, respectively. Those differences are mainly due to movements of the DBD (as for chain B) or the movement of one of the long helices (in chain D) (Fig. 35). The rmsd for common C_α atoms of the dimers CD/EF is 2.1 Å /2.0 Å compared to dimer AB. Therefore, the RovA-apo structure shows that the unliganded protein is relatively flexible, especially in the DBD.

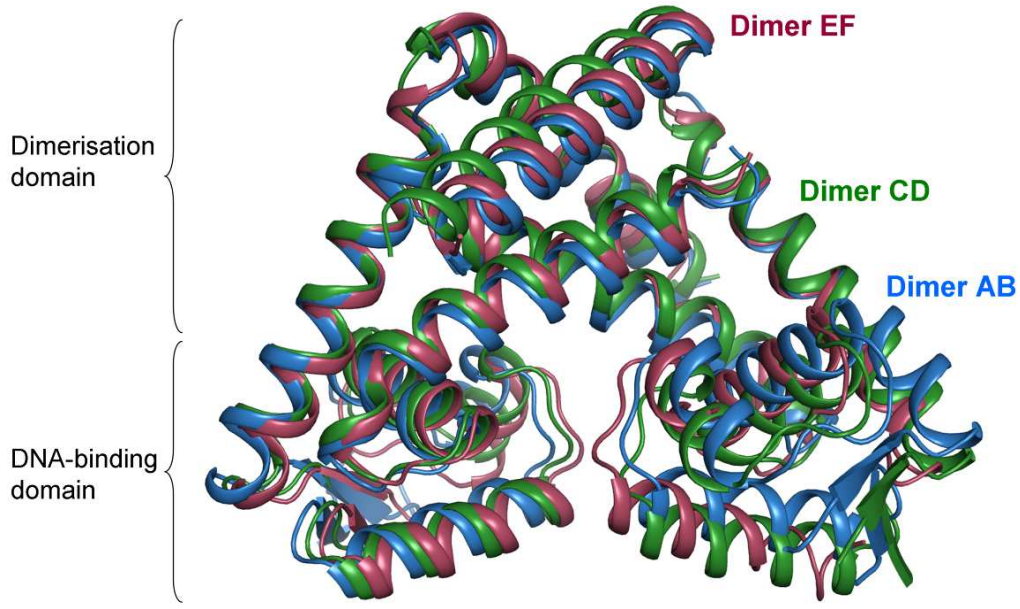


Fig. 35: Superpostion of the three RovA-apo dimers found in the asymmetric unit. The dimer AB is shown in blue, dimer CD in green and dimer EF in red.

Table 10: RovA apo and salicylate complex data collection and refinement statistics

	RovA apo	RovA + salicylate
X-ray source	ESRF ID14eh2	CuK_α Rigaku MicroMax 007HF
Space group	$P2_12_12_1$	C2
Unit cell dimensions *a*, *b*, *c* (Å)	67.2, 74.8, 181.1	88.8, 88.8, 67.7
Angles α, β, γ (°)	90, 90, 90	90, 95.6, 90
Wavelength (Å)	0.933	1.54
Resolution (Å)*	90.6 - 2.4 (2.46-2.4)	37.1 – 2.1 (2.2-2.1)
Measured	217609 (16046)	54970 (6989)
Unique reflections	36510 (2621)	15521 (2007)
I/σ_I*	16.79 (2.9)	14.9 (3.8)
Completeness (%)	99.8 (100)	98 (98.4)
Redundancy	5.9 (6.1)	3.5 (3.5)
R_{merge} (%)	7.3 (68.0)	5.7 (44.5)
Wilson B factor ($Å^2$)	57.5	48.2
Solvent content (%)	45	38.4
R_{work} (%)	22.8	31.4
R_{free} (%)	28.4	35.5
Protein atoms	6245	1998
Solvent atoms	98	73
R.m.s.d. from ideal geometry:		
-Bond lengths (Å)	0.017	0.022
-Bond angles (°)	1.71	2.19
Average B-factor		
-Protein (chain A/B)	24.63, 22.7	30.1, 22.8
-Protein (chain C/D)	33.7, 29.1	-
-Protein (chain E/F)	26.3, 55.2	-
-Solvent	29.5	32.5
Ramachandran plot (%):		
-Favoured region	97.3	88.9
-Allowed region	2.3	7.9

4A.1.16 The structure of RovA in complex with salicylate

The structure of RovA in complex with salicylate could be solved by molecular replacement using the homologous structure from SlyA from *S. typhimurium* (pdb code 3DEU). Even though the structure could not yet be completed and the R factors are still quite high (probably due problems with twinning), the electron density allowed to build a model with good stereochemistry. Two monomers were found in the asymmetric unit, forming a dimer. Similar to the RovA-apo structure, the wing of the DBD was not visible in the structure. The two monomers align well with an rmsd for common C_α atoms of 0.7 Å. In every monomer two salicylate molecules (Sal1 and Sal2) were bound (Fig. 36A).

Sal1 is bound in a pocket at the dimerisation interface between helices α1 and α5 of one monomer and α1 of the other monomer (Fig. 36B). This pocket is lined with hydrophobic residues such as W16, W34 and I107. The hydroxy-group of the salicylate molecule is bound by S7, while the negatively charged carboxyl group estiblishes hydrogen bonds to two arginine residues (R14 and R17). Therefore, this molecule is tightly bound within this pocket by hydrophobic and hydrophilic interactions. Sal2, however, is only weakly bound on the surface of the DBD (Fig. 36C). It is only held in place by hydrophobic interactions with residues such as I56 and I58 on a loop between helices α3 and α4. Thus, it is possible that the second salicylate molecule is only seen due to the unphysiologically high concentrations of salicylate (50mM) in the crystallization drop. However, the pocket of the first salicylate molecule is likely to be responsible for the ligand binding observed in vitro.

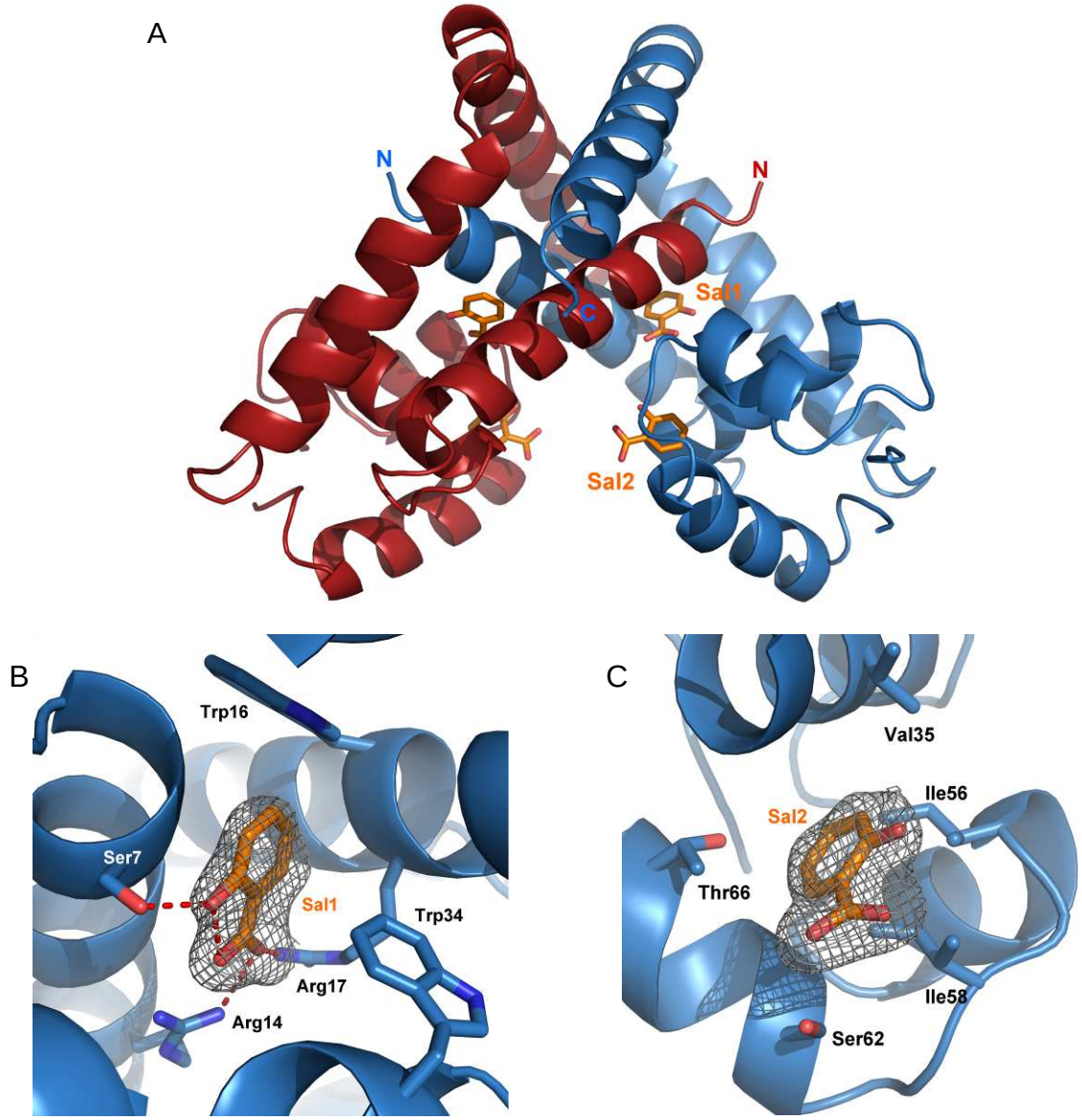


Fig. 36: Structure of RovA in complex with salicylate. A: Two salicylate (Sal1 and Sal2) molecules are bound to each RovA monomer. B: Sal1 is bound in a hydrophobic pocket. Electron density (2FO-FC) is shown as grey mesh at 1 sigma. C: Sal2 is bound at the surface of the DBD of RovA

4A.2 The transcription regulator RovM

4A.2.1 Purification of full-length RovM

The purification and crystallization of RovM and RovM-EBD were performed by Marieke Dieckmann as part of her bachelor thesis. The plasmid for the production of full-length RovM, pAKH43, was kindly provided by the group of Petra Dersch (HZI, Braunschweig). The plasmid was transformed into *E. coli* BL21 cells and a 2L culture producing the protein was grown. After lysis of the cells the protein in the supernatant was purified by Ni-NTA affinity chromatography. The column was washed to remove unspecifically bound proteins and RovM was eluted by a buffer with a high imidazole concentration. The elution fractions were tested on an SDS gel for purity and protein yield (Fig. 37).

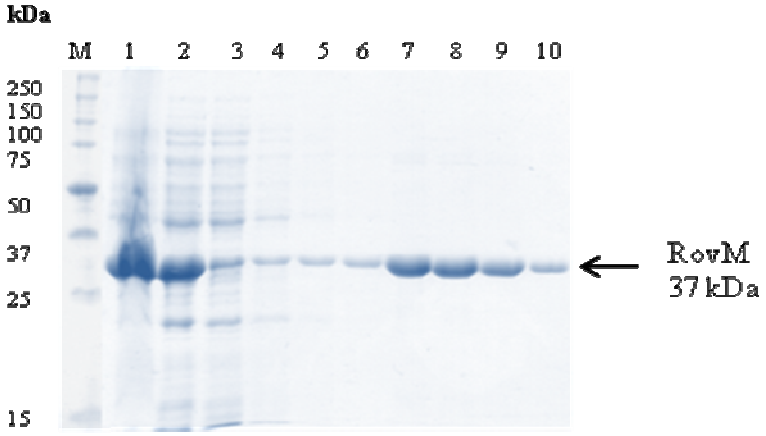


Fig. 37: SDS-gel of full-length RovM fractions after Ni-NTA purification. M: Marker, 1: pellet, 2: supernatant, 3: flow through, 4-6 waash, 7-10: elution.

To find a suitable buffer for gel filtration a ThermoFluor screen was performed. The screen showed that RovM was most stable at buffers with high pH and salt. Thus, 20mM Tris/HCl pH8, 100mM NaCl and 5mM DTT was chosen for gel filtration. The protein was concentrated to 1ml and applied to a 16/60 SD200 gel filtration column. The protein eluted as a single peak at 87ml (Fig. 38). This corresponds to a molecular mass of 174.9kDa while the calculated molecular mass of a monomer is 37kDa. Thus, the calculated oligomeric state of the protein is 4.7, indicating a tetramer in solution which has been observed for several other LTTR proteins as well (Muraoka *et al.* 2003).

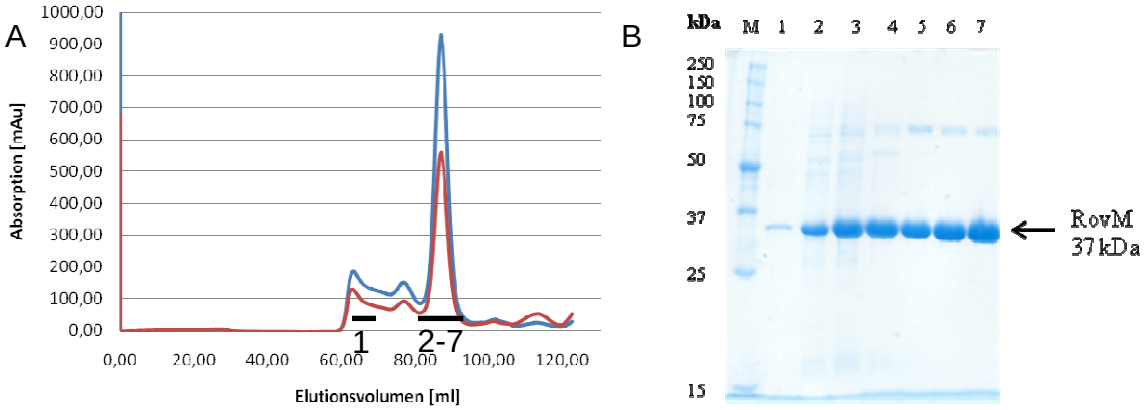


Fig. 38: Gel filtration of full-length RovM. A: Chromatogram of gel filtration run. Absorption at 280nm is shown in blue, absorption at 260nm in red. B: SDS gel of eluted fractions. M: Marker, 1: first small peak, 2-7 large peak.

4A.2.2 Design of the RovM-EBD construct

The domain boundary of the effector binding domain of RovM (RovM-EBD) was determined by comparison with structures and sequences of other LTTR proteins. The construct containing amino acids 92-310 was amplified from the plasmid pAKH43 which was used for the production of full-length RovM. The insert had the same restriction sites (NdeI and XhoI) as were used for the generation of pAKH43 and was cloned into the same plasmid.

4A.2.3 Purification of RovM-EBD

The protein RovM-EBD was produced and purified similar to the full-length protein. It was produced in BL21 cells, purified using Ni-NTA affinity chromatography and elution frations were analysed by SDS PAGE (Fig. 39).

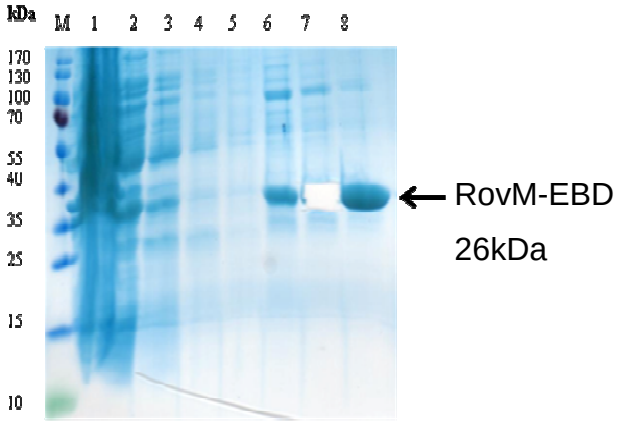


Fig. 39: SDS gel of RovM-EBD Ni-NTA chromatography. M = Marker, 1 = pellet, 2 = flow through, 3 – 5 = wash, 6 – 8 = elution fractions.

As a final purification step, gel filtration with the same buffer as for full-length RovM was done (Fig. 40). The protein eluted at 80 ml, which corresponds to a molecular mass of 57 kDa. With a calculated molecular mass of 26 kDa this results in an oligomeric state of 2.2, indicating a dimer in solution. This agrees with previous results that EBDs are dimers in solution (Ezezika *et al.* 2007).

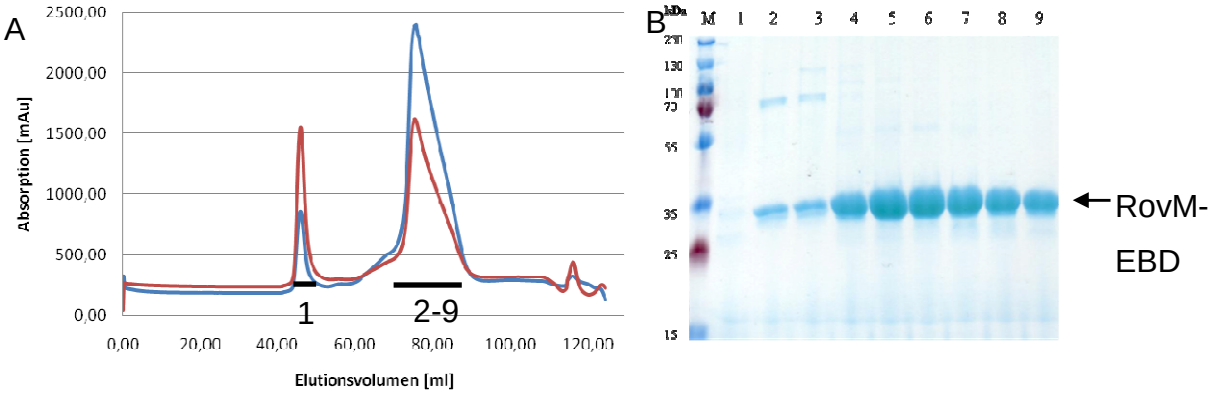


Fig. 40: Gel filtration of RovM-EBD. A: Chromatogram of gel filtration run. Absorption at 280nm is shown in blue, absorption at 260nm in red. B: SDS gel of eluted fractions. M: Marker, 1: first small peak, 2-9 large peak.

4A.2.4 Crystallization of full-length RovM and RovM-EBD

To find initial crystallization conditions, several commercially available crystallization screens of full-length RovM and RovM-EBD were set up. Despite extensive screening, no crystals of full-length RovM could be obtained. RovM-EBD on the other hand crystallized under several conditions with a pH of about 7 and higher molecular mass PEG (Fig. 41).

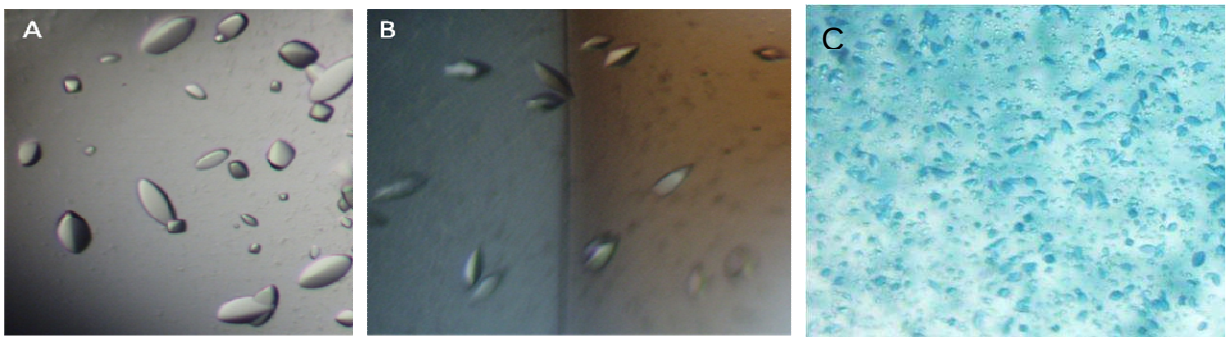


Fig. 41: Initial RovM-EBD crystals. A: 0,1 M KCl; 0,1 M HEPES pH 7,0; 15 % PEG 6000 B: 0,1 M HEPES pH 7,5; 20 % PEG 8000. C: Crystals stained with methyleneblue (Izit).

Optimization of the crystallization conditions to 0.1 M HEPES pH 8.0 and 12.5 % PEG 6000 yielded large crystals that grew within 3 weeks (Fig. 42). To obtain phases, RovM-EBD crystals were soaked with 50mM $SmCl_3$ for 10min or 400mM magic triangle, both in reservoir solution. Before measurement the crystals were transferred into a drop with 0.1 M HEPES pH 8.0, 25 % PEG 6000 and 10% glycerol for cryo protection and then flash frozen.

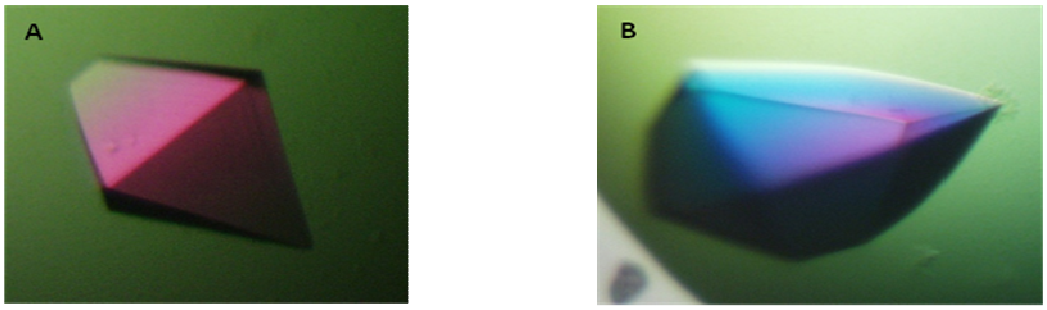


Fig. 42: Optimised RovM-EBD crystals used for data collection. 0.1 M HEPES pH 7.5 (A) or pH 8.0 (B) and 12.5 % PEG 6000.

4A.2.5 Data collection

The RovM-EBD crystals belonged to the tetragonal space group $I4_122$ with unit cell parameters a=b= 69.47 Å and c=351.22 Å. Calculation of the Matthews coefficient to V_M = 2.12 $Å^3$ Da^{-1} indicated two monomers per asymmetric unit corresponding to a solvent content of 43%. To solve the phase problem, molecular replacement was tried using several EBD structures (1AL3 (CysB), 2FYI (Cbl), 2H98 (CatM) and 2H9B (BenM)) as well as parts of these, but no solution could be obtained. Therefore, the crystals soaked with $SmCl_3$ and magic triangle were measured at the home-source at

1.54 Å. This is not the peak wavelength for samarium (L-I edge: 1.60 Å) or iodine (L-I edge: 2.39 Å), as incorporated in the magic triangle, but rather a high remote, yet, sufficient high anomalous signal was obtainable (Fig. 43). In particular, the Sm L-I edge was quite close giving a strong signal with a calculated f´´ of 12.2e and an f´ of -5.5e.

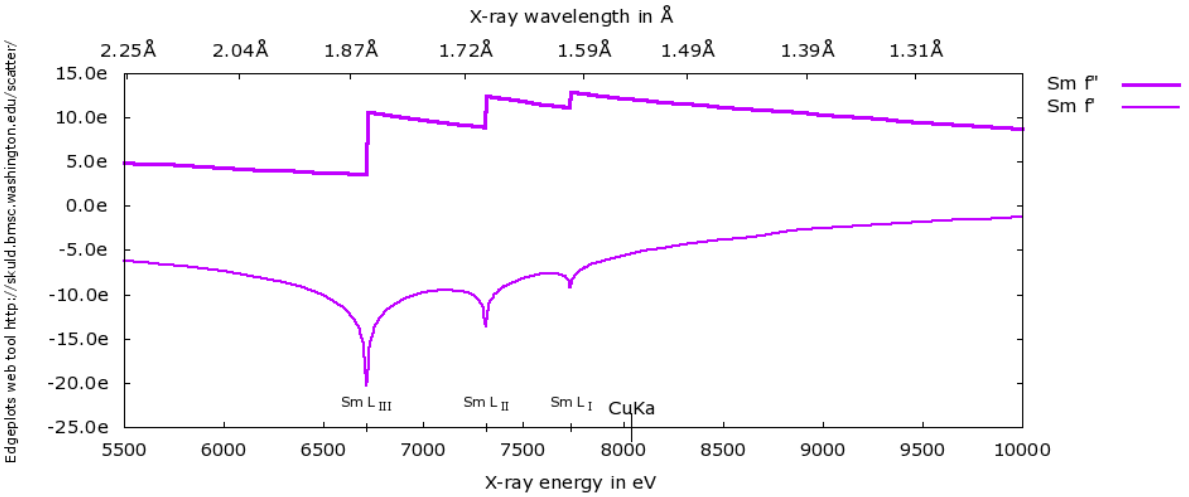


Fig. 43: Plot of the anomalous scattering contributions f´ and f´´ for samarium. f´´ is shown as a thick line, while f´´ is shown as the thin line. The absorption edges and the Cu K_α wavelegh are marked. Generated by the edge plot tool from http://skuld.bmsc.washington.edu/scatter/AS_form.html. (Evans & Pettifer 2001)

Substructure solution with ShelX (Sheldrick 2008) was successful for both derivatives finding 3 iodine sites in the magic triangle derivative (corresponding to one magic triangle molecule) and for the Sm derivative 4 strong and several weak sites were found. Yet, the one bound magic triangle molecule did not generate enough phasing power to obtain useful phases. The Sm derivative on the other hand provided useful phases after refinement of the heavy atom positions by Phaser (McCoy *et al.* 2007) and density modification by Parrot (Cowtan 2010). A first model was built by Buccaneer (Cowtan 2006) and then refined manually in Coot (Emsley *et al.* 2010) and with Refmac5 (Murshudov *et al.* 1997). After several rounds of refinement and a last TLS refinement step, R_{work} and R_{free} reached 22.1% and 29.5%, respectively. The rather large difference between these two values probably results from the bad electron density for some parts of the B chain. The final model was validated by Molprobity (Chen *et al.* 2010) and the Ramachandran plot showed 97.8% of the residues in the favoured region and 2.2% in the allowed region.

Table 11: RovM-EBD data collection and refinement statistics

	SAD	native
X-ray source	CuK_α Rigaku MicroMax 007HF	ESRF ID 29
Space group	$I4_122$	$I4_122$
Unit cell dimensions *a*, *b*, *c* (Å)	69.03, 69.03, 352.49	69.47, 69.47, 351.22
Wavelength (Å)	1.54	0.97
Resolution (Å)*	88.6-2.5	87.8-2.4 (2.53 - 2.4)
Measured reflections*	192810 (23819)	182593 (22862)
Unique reflections*	26240 (4108)	17557 (2485)
I/σ_I*	28.7 (4.9)	16.4 (3.5)
Completeness (%)*	99.4 (96.7)	99.9 (100)
Redundancy*	7.3 (5.8)	10.4 (9.2)
R_{merge} (%)*	6.3 (41.6)	7.4 (54.0)
Wilson B factor ($Å^2$)	33.75	59.3
Solvent content (%)	43%	43%
R_{work} (%)		22.1 (23.4)
R_{free} (%)		29.5 (30.2)
Protein atoms		2142
Solvent atoms		142
r.m.s.d. from ideal geometry:		
Bond lengths (Å)		0.014
Bond angles (°)		1.575
Average B-factor ($Å^2$):		
Protein (chain A)		56.4
Protein (chain B)		62.7
Solvent		55.3
Ramachandran plot (%):		
Favoured region		97.8
Allowed region		2.2

4A.2.6 Crystal structure of RovM-EBD

The structure of RovM-EBD consists of two distinct domains EBDI and EBDII, which are connected by two crossovers at the two β-strands β4 and β10. EBDI is composed of a 5 stranded β-sheet and 3 α-helices around the sheet, while EBDII consists of a strongly twisted 5 stranded β-sheet with four α-helices around it (Fig. 44A and B). The asymmetric unit contains two chains of RovM-EBD. These adopt the same fold, yet, due to crystal packing, show slight differences (Fig. 44C). For chain A

all residues from 99 to 289 could be modeled except for residue 159, whereas for chain B only residues 100 to 280 with a gap at residue 128 were built. Most of the differences are in the EBDI while the EBDII is nearly identical. For example, the C-terminal α-helix is well resolved in chain A while the density was much worse in chain B, where it is missing some crystal contacts. These differences are signified by the r.m.s.d. of 0.91 Å for all common C_α atoms.

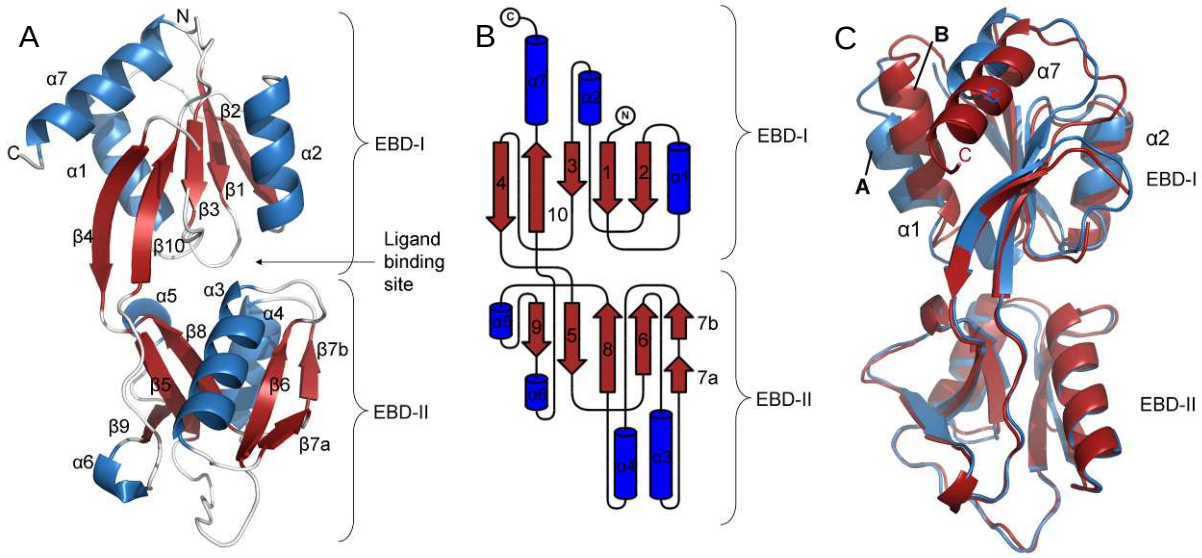


Fig. 44: Structure of RovM-EBD monomer. A: Cartoon representation showing the two distinct domains and the possible ligand binding site; B: topology plot, generated by Topdraw (Bond 2003); C: Superposition of chain A and B. Chain A is shown in blue, chain B in red.

4A.2.7 Oligomeric state of RovM-EBD

Full length LTTRs can exist in several oligomeric states, ranging from dimers over tetramers to octamers (Monferrer *et al.* 2010). The gel filtration analysis of full length RovM indicated a tetrameric assembly, similar to CbnR (Muraoka *et al.* 2003). EBDs on the other hand have been reported to be dimers in solution (Ezezika *et al.* 2007). The gel filtration analysis of RovM-EBD indicates that RovM-EBD is also a dimer in solution. As a matter of fact, in the asymmetric unit there are two monomers present which share a quite large interface of 845.7 $Å^2$ (Fig. 45). This interaction is mostly mediated by helix 1, sheet 4 and sheet 9 as well as the loop preceeding sheet 9. Yet, comparison with structures of other LTTR EBDs showed that this interface is different from the usual dimer interface. Thus, the crystal contacts with the symmetry mates of RovM-EBD were evaluated using the PISA server and it became apparent, that RovM-EBD shows several different significant interfaces within the crystal packing.

The strongest of these is the interface between chain B and its symmetry mate B´ (Fig. 45). This interface is the same as the typical EBD dimer interface, where the two monomers are arranged in a head-to-tail fashion with the EBD-I of one molecule interacting via helix 1 with the helix 4 of the EBD-II of the other molecule covering interface area of 883.8 $Å^2$. Yet, the interface between A and its symmetry mate A` is different (Fig. 45). The two monomers are shifted relative to each other and thus the interaction is mediated by helices 1 and 4 and sheet 2. Even though the interface area is, with 945.6 $Å^2$, larger than the other interfaces, it is less stable and was considered to be less relevant by the PISA server. From these data we conclude that RovM-EBD probably is a dimer in solution characterised by the BB´ interface, while the strong interaction between chain A and B in the crystal leads to a distortion of the AA` interface. Another weak interface was observed between two A or B chains (termed AA`` or BB``) (Fig. 45). This interaction covers an area of 681.8 $Å^2$ and 642.7 $Å^2$ for AA`` and BB``, respectively and is mediated by helices 2 and 3.

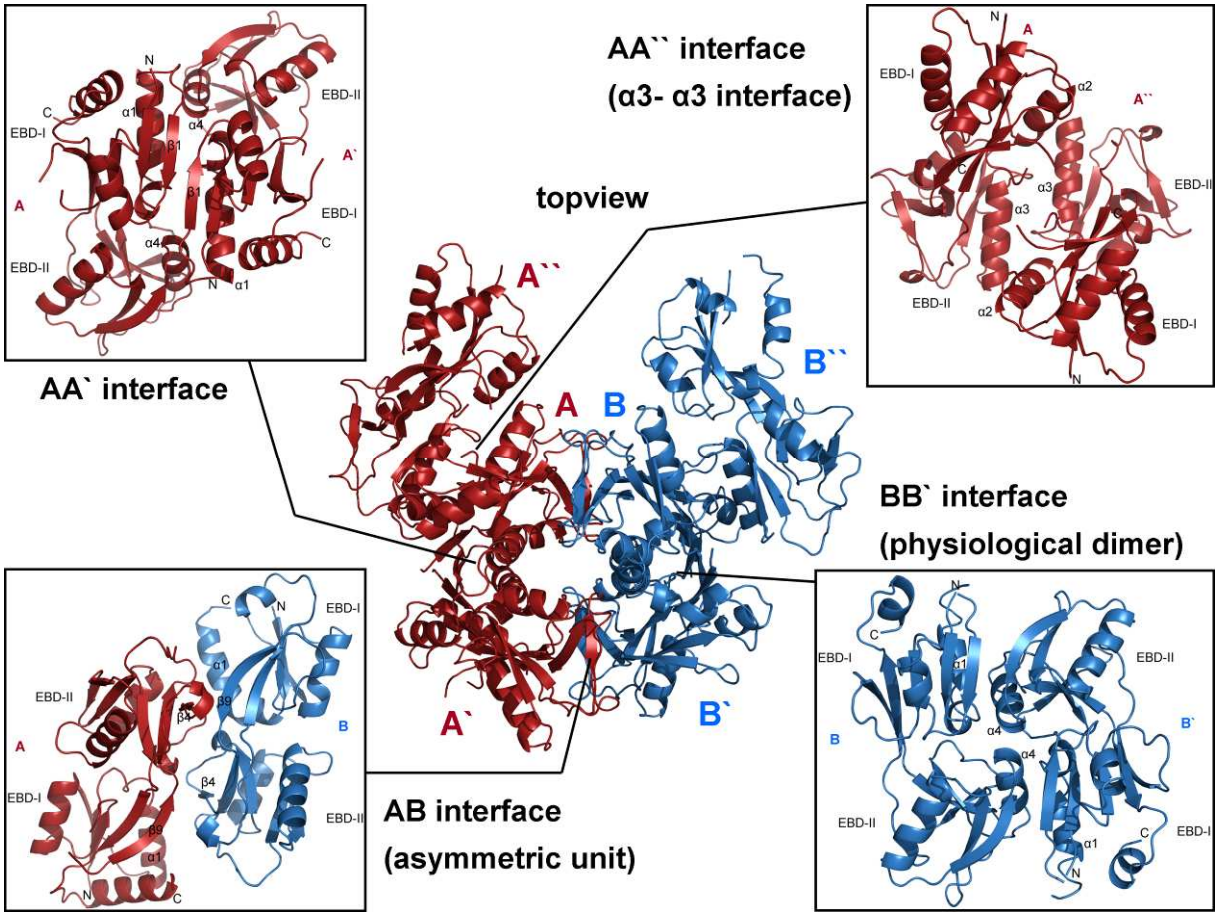


Fig. 45: Structures of the interfaces between RovM-EBD monomers in the crystal. Chain A is shown in blue, chain B in red. The AB interface is seen in the asymmetric unit. The BB` interface is seen in all structures of LTTRs and is thus termed the physiological dimer. The AA` interface is similar to the BB` interface but the two EBDs are shifted relative to each other. The AA`` interface (and the BB`` interface, not enlarged) is seen in some other LTTRs and seems to mediate tetramer formation.

4A.2.8 Ligand binding pocket

Small molecules, called inducers, that bind to LTTR proteins have been shown to play an important role in gene regulation (Schell 1993). Often these molecules are related to the genes they influence, such as benzoate in aromatic compound degradation (Ezezika *et al.* 2007). They usually bind in a cavity between the two domains (EBDI and EBDII) of the EBD, which is thought to result in a conformational change in the EBD. This is translated via the long lever-like linker helix into a repositioning of the DBD and a resulting relieved bending of the DNA (Schell 1993). So far, no inducer is known for RovM, but the EBD structure has a cavity (roughly 10.5 Å wide, 7.1 Å high and 6.5 Å deep) in the interface between the two domains that is of a suitable size to bind a small ligand. The cavity is lined with two

hydrophobic residues (Y198 and L222) and several hydrophilic residues (D106, D107, S137, T155 and R238), suggesting that a rather hydrophilic inducer molecule could bind in this cavity.

4B Polyketide synthases

4B.1 The octenoyl-CoA carboxylase/reductase CinF

4B.1.1 Purification of CinF

The cinF gene from *S. coelicolor* was cloned and the protein produced with an N-terminal SUMO-tag in the group of Rolf Müller (Saarland University, Saarbrücken). The protein was purified by affinity chromatography and the tag was cleaved off (Fig. 46). Then the protein was kindly provided for further experiments.

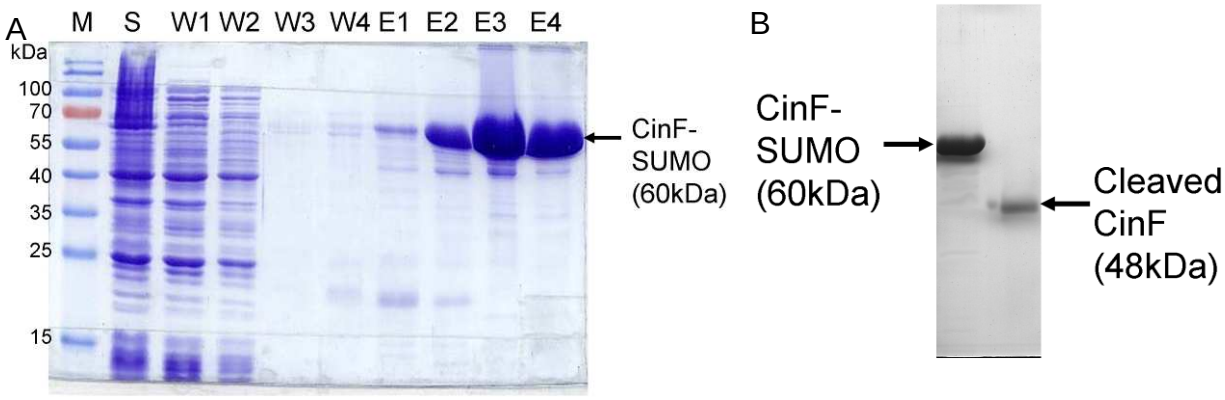


Fig. 46: SDS gels of CinF (kindly provided by Rashid Shwan, Saarland University, Saarbrücken). A: M: marker; S: soluble lysate; W1-4: wash; E1-4: elution fractions. B: CinF before and after cleavage of the SUMO tag.

First, a ThermoFluor screen was performed to find a buffer in which the protein was stable. The screen showed that CinF is most stable at higher pH values and thus 10mM Tris pH 8 and 100mM NaCl was used as buffer for CinF. Subsequently, the protein was further purified by gel filtration on a 16/60 gel filtration column using the gel filtration buffer (Fig. 47A). The elution fractions were analysed by SDS PAGE and pure fractions were pooled and concentrated for crystallization (Fig. 47B).

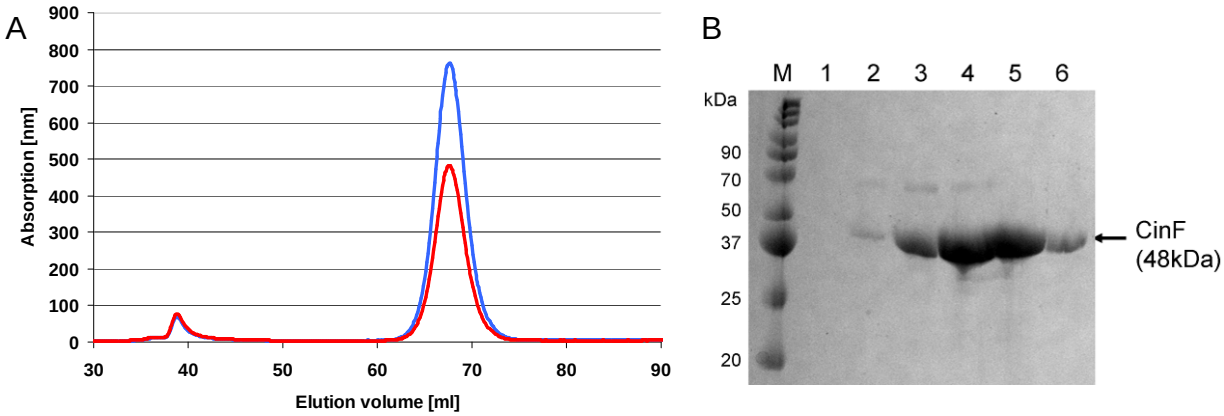


Fig. 47: Gel filtration of CinF. A: Elution profile. Absorption at 280nm is shown in blue, absorption at 260nm in red. B: SDS gel of eluted fractions. M: Marker, 1-6 large peak (elution volume 62-72ml).

4B.1.2 Crystallization of CinF

Initial screening for crystallization conditions yielded small plate-like crystals under the condition: 0.1M CHES pH 9.5 and 10% PEG3000 (Fig. 48A and B). These could only be reproduced by streak seeding from the initial crystals into fresh crystallization drops. After optimization of the crystallization conditions in 24 well plates, larger crystals were obtained that were still very thin and fragile and did not diffract very well. Seeding from these into the drops of the initial screens that had remained clear yielded several new crystallization conditions. Of these, 0.1M Tris pH 7, 20% glycerol and 12% PEG8000 produced thicker crystals and was already a cryo solution reducing damage to the fragile crystals by handling. The protein was crystallized in drops containing 1µl protein and 1µl reservoir buffer that were incubated for one day and then streak seeded from older crystals. Large crystals grew within one day and reached their final size within 3 days (Fig. 48C). Crystals containing $NADP^+$ and octenoyl-CoA were produced by adding 10mM NADP and 10mM octenoyl-CoA to the protein solution before crystallization.

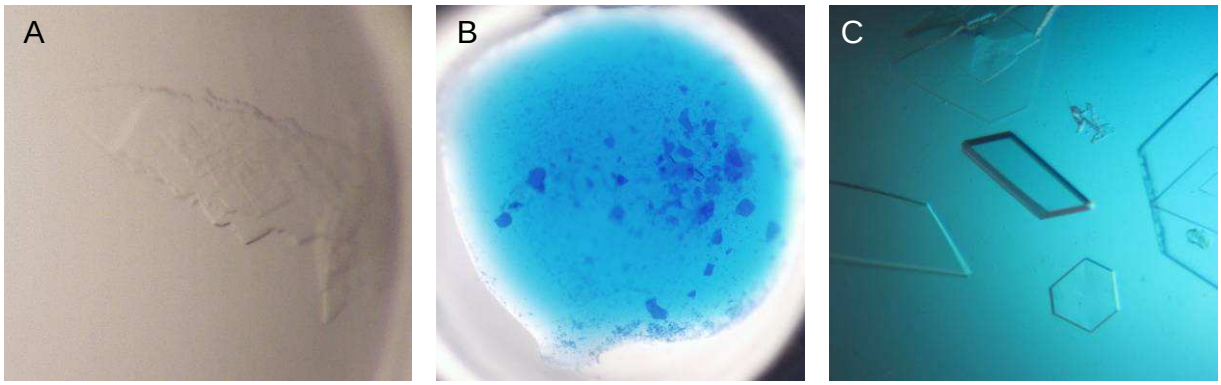


Fig. 48: CinF crystals. A: Initial crystal in 0.1M CHES pH 9.5 and 10% PEG3000; B: Initial crystals stained with methyleneblue (Izit); C: Optimized crystals in 0.1M Tris pH 7, 20% glycerol and 12% PEG8000

4B.1.3 Data collection

A native CinF dataset was measured at ESRF (Grenoble, France) beamline ID 29, equipped with an ADSC Q315R detector. A Micromax-007 HF rotating copper anode X-ray generator (Rigaku) with a Saturn 944+ CCD detector (Rigaku) was used to measure the data set of CinF with NADP and octenoyl-CoA diffracting to 1.9 Å. The CinF crystals belong to the space group $P2_1$ with unit cell dimensions a= 96.01 Å, b= 83.30 Å and c= 122.74 Å and α=γ= 90° and β= 110.96°. The data processing was carried out using XDS (Kabsch 2010) and scaling was done with XSCALE. The Matthews coefficient (Matthews 1968) of V_M = 2.39 $Å^3$ Da^{-1} indicated four monomers per asymmetric unit corresponding to a solvent content of 48%. To obtain initial phases molecular replacement by MOLREP (Murshudov *et al.* 1997) was perfomed using the putative crotonyl-CoA carboxylase/reductase (3KRT) as a search model, finding four monomers in the asymmetric unit. The structure was refined with REFMAC5 (Murshudov *et al.* 1997) and completed in Coot (Emsley *et al.* 2010). For the data set of CinF in complex with its substrates the model of the apo-structure was used for molecular replacement. Clear density for NADP was visible in all monomers and also for octenoyl-CoA in two monomers, whereas in the other two monomers the density was much weaker. Structure and restraint files for octenoyl-CoA were generated using the Dundee Prodrg server (Schuttelkopf & van Aalten 2004). The octenoyl-CoA was build into the monomer with the best density (chain A) and, by applying the symmetry, placed in the other monomers as well. Water molecules were added manually and using the water find tool in Coot. A final TLS refinement (Painter & Merritt 2006a) step using five TLS groups per monomer as determined by the TLS

Motion Determination Server (Painter & Merritt 2006b) was carried out with REFMAC5. The final R values were 21.8 / 27.6 % (Rwork / Rfree) for the apo-form and 19.9 / 24.6 % for the ligand bound form, respectively.

In Table 12 data collection and refinement statistics are listed. MolProbity (Chen *et al.* 2010; Davis *et al.* 2007) validation indicated a good quality model. Nearly all residues fall within the allowed regions of the Ramachandran plot (Ramachandran & Sasisekharan 1968). Only in the complex structure H313, which is involved in substrate binding, is an outlier in all chains. The relevance of the crystal contacts was assessed using the PISA server (Krissinel & Henrick 2007).

4B.1.4 CinF structure

In this work we present the structure of the octenoyl-CoA carboxylase/reductase CinF in its apo-form at 2.0 Å resolution as well as in complex with its ligands NADP and octenoyl-CoA at 1.9 Å resolution. The asymmetric unit contains four CinF monomers (chain A-D) (Fig. 49A). The apo structure had weak electron density in some parts, probably due to flexibility in the absence of the ligands. The residues 1-347 were modelled for chains A and B with gaps between residues 275-286 and 335-347 and residues 1-352 were built for chains C and D with gaps between 275-286, 313-317 and 335-347. These parts were well defined in the structure with NADP and octenoyl-CoA, which comprises the residues 1-445. Apart from the sections missing in the apo-structure, both structures align quite well with an average rmsd of 0.63 for shared C_α atoms for each chain. Thus, unless otherwise stated, it is referred to the structure of the complex in the rest of the text.

The monomers superimpose very well upon each other with an rmsd of 0.13 Å between chain A and chain B as well as 0.25 Å and 0.21 Å between chain A and chain C or chain D, respectively. Each CinF monomer can be divided into two domains: the N-terminal catalytic domain (residues 1-201 and 361-445), which is necessary for substrate binding and catalysis, and the C-terminal cofactor binding domain (residues 202-360), which binds NADP(H) (Fig. 49B). The cofactor binding domain exhibits a Rossmann fold, containing two mononucleotide binding βαβαβ motifs and an additional strand β14 (Fig. 49C). The catalytic domain consists of a seven stranded antiparallel β-sheet, a separate three stranded antiparallel β-sheet

and 11 α-helices. The long α-helix α9, which connects both domains, is considered to be part of both domains as the N-terminal part of the helix is involved in substrate binding, whereas the C-terminal part is involved in NADP(H) binding.

Table 12: CinF data collection and refinement statistics

	apo	octenoyl-CoA/NADP complex
Beamline	ESRF ID 29	CuK_α Rigaku MicroMax 007HF
Space group	$P2_1$	$P2_1$
Unit cell dimensions *a*, *b*, *c* (Å)	101.64, 85.22, 113.97	96.01, 83.32, 122.74
α,β,γ (°)	90, 107.4, 90	90, 110.9, 90
Wavelength (Å)	0.976	1.54
Resolution (Å)	48.48-2.1 (2.25-2.1)	47.2-1.9 (2.0-1.9)
Measured reflections	357025 (63670)	505455 (61327)
Unique reflections	107076 (19855)	140724 (18977)
I/σ_I*	8.78 (3.68)	13.85 (3.6)
Completeness (%)	98.76 (98.83)	98.8 (94.1)
Redundancy	3.3 (3.2)	3.6 (3.2)
R_{merge} (%)	15.8 (62.2)	8.7 (42.6)
Wilson B factor ($Å^2$)	30.8	27.0
Solvent content (%)	48	48
R_{work} (%)	21.78	19.94
R_{free} (%)	27.65	24.65
Protein Atoms	12550	13389
Ligand Atoms	-	400
Solvent Atoms	1515	1905
r.m.s.d. from ideal geometry		
Bond lengths (Å)	0.0132	0.0106
Bond angles (°)	1.52	1.61
Average B-factor ($Å^2$)		
Protein (chains A, B, C,D)	5.66, 6.37, 9.45, 6.95	3.43, 4.29, 4.09, 4.87
NADP	-	15.3, 16.4, 25.2, 22.8
Oct-CoA	-	33.6, 39.1, 75.9, 55.78
Solvent	18.37	14.2
Ramachandran plot (%)		
favoured	95. 8	96.0
allowed	3.6	3.9

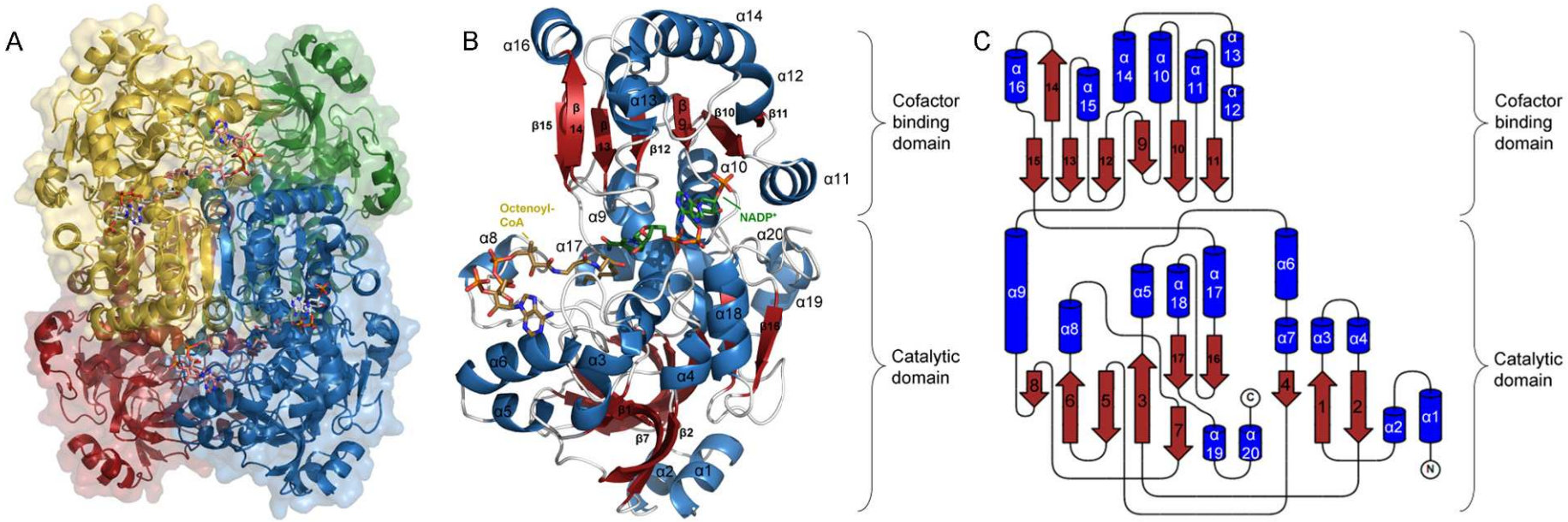


Fig. 49: Structure of CinF. A: Quarternary structure of CinF showing the dimer of dimers fold. The monomers are shown as cartoon and transparent surface in green, blue, yellow and red. The ligands are shown as sticks. B: Cartoon representation of CinF monomer. Helices are colored in blue, β-sheets in red and loops in white. The ligands are shown as sticks, the $NADP^+$ is green, the octenoyl-CoA yellow. C: Topology plot of CinF, showing the domain organisation and the secondary structure of CinF drawn with Topdraw (Bond 2003).

4B.1.5 Oligomerisation

In the asymmetric unit four CinF monomers were found in a tetrameric dimer of dimers assembly (Fig. 49A). Size exclusion chromatography of CinF suggests a molecular mass of 176 kDa (the calculated molecular mass of CinF is 48kDa) corresponding to a calculated oligomeric state of 3.7. Analysis of the most probable assembly using the PISA server also suggests a tetramer covering an area of 25,410 $Å^2$. The strongest interaction between two monomers is between the cofactor binding domains of chains A and D as well as chain B and C with an average interface area of 1475 $Å^2$ forming two dimers. This interaction is mainly mediated by β15 of both subunits, thus creating a continuous 12 stranded β-sheet, β14, interacting with the same β-strand from the other chain, and helix α16, which is inserted between the two domains of the other chain. Interestingly, β14 and the preceding loop are disordered in the apo-structure, probably due to the lack of stabilisation by the ligands. The two dimers interact via two interfaces to form the tetramer. The first tetramer interface is between the catalytic domains of chain A and B and chain C and D covering an average area of 1252 $Å^2$. Here helix α8 of one chain is wedged between the helices α8 and α17 of the other chain. The second tetramer interface is relatively weak covering an area of only 629 $Å^2$. It is mediated by loop regions between α9 and β9 as well as between α10 and β9 in the cofactor binding domain of chains A and C and chains B and D.

4B.1.6 $NADP^+$ binding

Between the two domains of CinF there was a strong density allowing unambiguous placement of the $NADP^+$ molecule in all four subunits (Fig. 50A). The $NADP^+$ molecule assumes an extended conformation and is only slightly bent around two loops of the protein, the loop preceding β14 and the loop between β12 and α15. These loops make several interactions with the $NADP^+$ molecule and are disordered in the apo-structure. The $NADP^+$ molecule is tightly bound to the protein via many hydrogen bonds and hydrophobic interactions. The adenine group is bound mostly via water mediated contacts and hydrophobic interactions with serine residues 338 and 339. The specificity of reductases for NAD(H) or NADP(H) is determined by the residues that bind to the 2`phosphate group of NADP(H) or prevent the placement of the phosphate group. In the CinF structure the 2`phosphate group is recognized by

hydrogen bonds with Lys257 and with the OH group and the backbone nitrogen of Ser253 as well as several water mediated interactions with e.g. Arg272 and Ser230. This shows that NADP(H) is indeed the cofactor and not NAD(H). The pyrophosphate group is tightly bound via hydrogen bonds with Gln405 and the backbone nitrogens of Gly232 and Leu233 as well as several water mediated interactions. The last two residues are located at the N-terminal end of helix α10, so that the positive dipole of the helix can compensate the negative charge of the phosphates These residues are part of a GXXGXXG motif, which has been shown to be the cofactor binding motif in quinone oxidorectuctases, with which CinF shares some homology (Edwards *et al.* 1996). The nicotinamide group, which assumes an *anti*-conformation, on the other hand is inserted in a hydrophobic pocket lined by Leu201, Thr205, Leu233 and Cys335. Only the amide group is held in position by hydrogen bonds with Thr205 and His361.

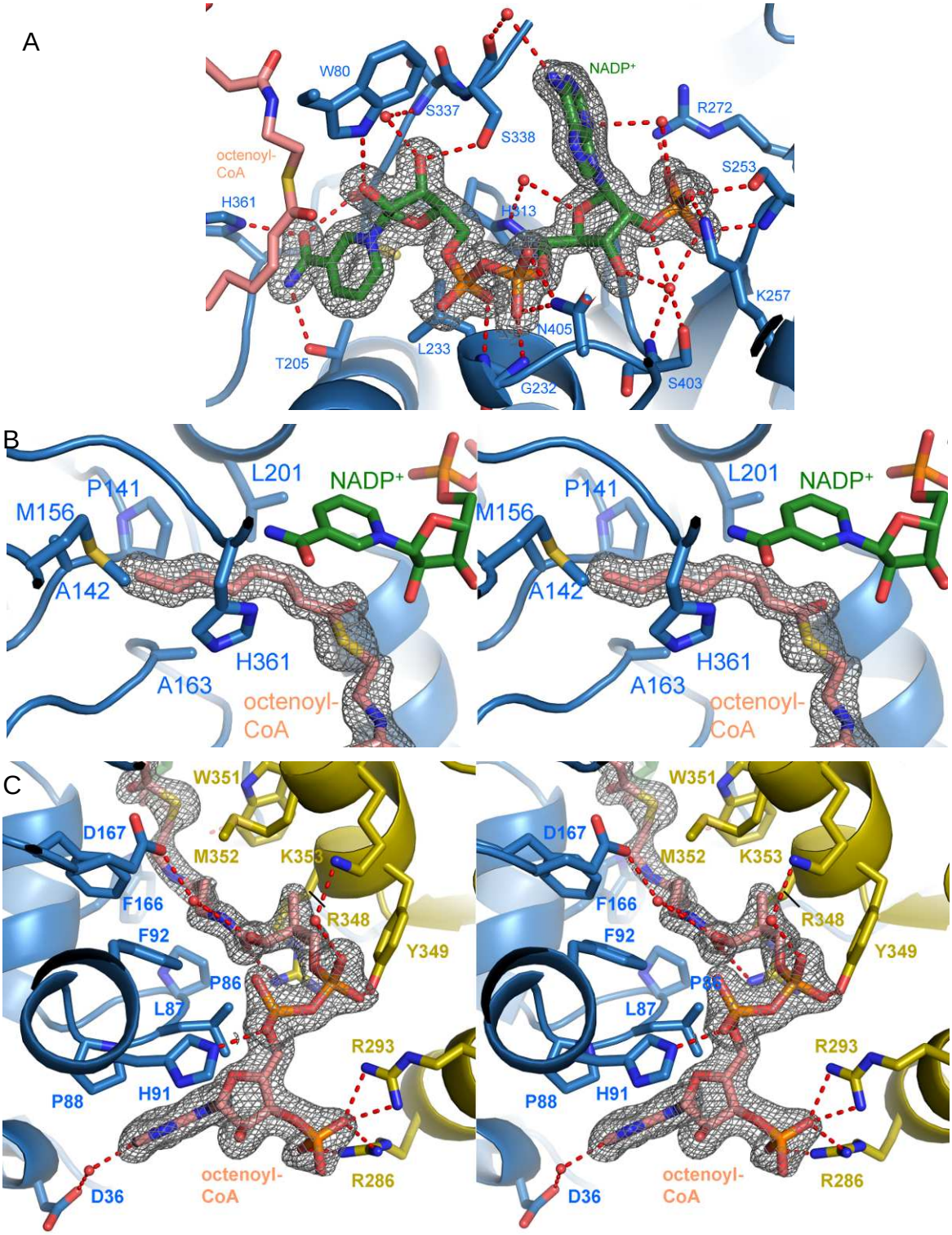


Fig. 50: Ligand binding by CinF. CinF is shown as cartoon and sticks, chain A is colored blue and chain D yellow. $NADP^+$ and Octenoyl-CoA are shown as green and pink sticks, respectively. The electron density is shown at 1.0 sigma as grey mesh. A: Interactions between $NADP^+$ and CinF showing the residues responsible for cofactor specificity. B: Stereo image of the hydrophobic substrate binding pocket for the octenoyl-chain. C: Stereo image of the binding of the CoA at the interface between chain A and D.

4B.1.7 Octenoyl-CoA binding

Good density for the octenoyl-CoA in the monomers A and B, whereas the density was poorer in monomers C and D, suggesting only partial occupancy of the ligand. Octenoyl-CoA was built into the monomer with the best density (A) and then placed in the other monomers by using the non-crystallographic symmetry. The binding of octenoyl-CoA is described here for chain A. The long and flexible ligand is bound in an extended conformation, in which the acyl-group is buried within the catalytic domain while the CoA is bound on the surface of the monomer and held in place by interactions with the dimer partner (chain D) of chain A (Fig. 50B). The acyl-group is accommodated by a hydrophobic pocket created by residues Pro141, Ala142, Met156, Ala163, Leu201, His361 and Gly362 as well as the nicotinamide group of the $NADP^+$ molecule. This places the C3 of the octenoyl-chain in a perfect position for the hydride transfer from an NADPH molecule.

The pantethein part of the CoA is bound in the interface between the catalytic domain of chain A and α14 and α16 of chain D (Fig. 50C). It is mostly contacted by hydrophobic interactions with Trp80, Pro86, Leu87, Phe92 and Phe166 from chain A as well as Tyr349 and Met352 from chain D. Additionally, there is a hydrogen bond from Arg348 and some water mediated interactions. The rest of the CoA molecule winds around helix α6 and the preceding loop. The phosphates of the CoA molecule are contacted by His91 from chain A and Arg286, Arg293 from α14 and Tyr349 from α16 of chain D. These residues are not visible in the apo-structure and are probably flexible without a ligand bound.

The ribose and the adenine parts of the CoA molecule on the other hand are positioned on the surface of monomer A. They are only held in place by hydrophobic interactions with Leu87, Pro88 and His91 and a water bridge to Asp36.

4B.1.8 CO_2 binding

It has been shown that CCRs use CO_2 to carboxylate their substrate from the *re* face, in an *anti* position to the hydride transfer (Erb *et al.* 2009). We aimed to confirm this by solving the structure of CinF in complex with CO_2. However, cocrystallization with as much as 100mM Na_2CO_3 or CS_2 as a CO_2 mimic yielded no visible density for CO_2

(or CS_2) in the structure. Thus, we tried to aerate the crystals with up to 40bar CO_2 in a pressure chamber for 15min directly before freezing, yet even this measure was unsuccesful. Probably, the affinity for CO_2 is low and it is thus not visible in the structure. Therefore, we docked the CO_2 molecule into the structure *in silico* by GOLD (Verdonk *et al.* 2005). Several very similar solutions were found, all of them with a relatively low affinity (Fig. 51). The CO_2 molecule is held in place by hydrogen bonds with residues N77 and E167 as well as hydrophobic interactions with the underlying F166. It is in an anti position relative to the NADP and the distance is about 3.3Å from the C-atom of the CO_2 to the C2. Therefore, we deem this a very probable orientation as the CO_2 is optimally positioned for a nucleophilic attack from C2 of the octenoyl-CoA.

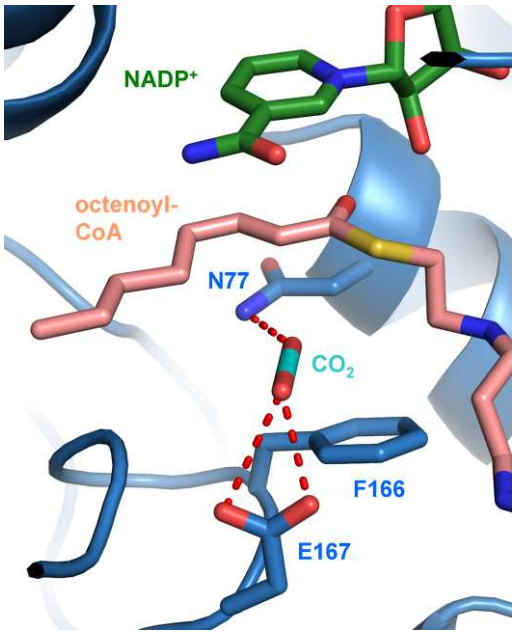


Fig. 51: CO_2 docked into the structure of CinF. CO_2 is shown as cyan sticks, CinF is shown in blue, octenoyl-CoA in pink and NADP in green.

4B.2 The thioesterase SpirTE

4B.2.1 Production and purification of SpirTE

The purification and initial crystallisation were performed by Christoph Bürth as part of his diploma thesis, crystallization of the Se-Met labelled protein was performed by Marieke Dieckmann. The plasmid carrying the SpirTE gene was kindly provided by Kathrin Buntin in the group of Rolf Müller (Saarland University, Saarbrücken). The protein was produced in BL21 cells from 2L LB medium. The cell lysate was applied to a Ni-NTA column and SpirTE was eluted by buffer containing 250mM imidazole. The elution fractions were tested for purity by SDS PAGE (Fig. 52).

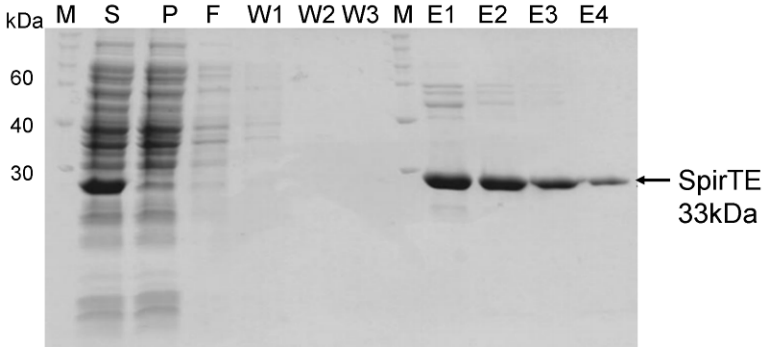


Fig. 52: SDS gel of SpirTE Ni-NTA chromatography. M: marker; S: soluble lysate; P: insoluble lysate; F: flow through; W1-3: wash; E1-4: elution fractions

The protein was further purified by gel filtration using a 16/60 SD200 column (GE healthcare). The protein eluted as a single peak and the corresponding fractions were analysed by SDS PAGE (Fig. 53). The protein elutes at about 75ml, which corresponds to a molecular mass of about 80kDa. Considering the calculated molecular mass of a SpirTE monomer of 33kDa this results in a calculated oligomeric state of 2.4, indicating a dimer, similar to other PKS thioesterases (Tsai *et al.* 2001).

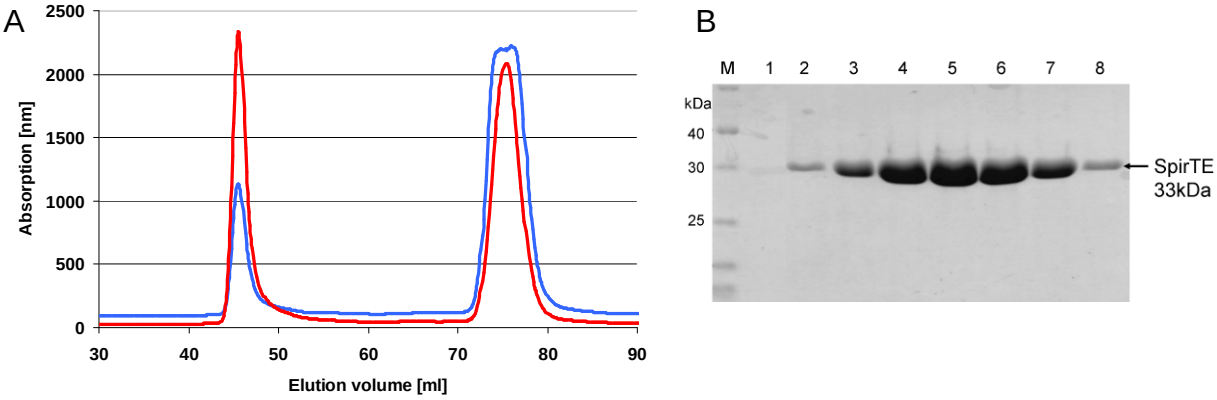


Fig. 53: SpirTE gel filtration. A: Chromatogram. B: SDS gel. M: marker; 1-8 elution fractions (elution volume 70ml - 82ml).

The total yield of pure SpirTE from 1L culture was 30mg. After gel filtration, the protein was concentrated to 7mg/ml for crystallization.

4B.2.2 Crystallization of SpirTE

Initial crystallization screens resulted in several hits under conditions with pH around 5 and PEG as precipitant. The crystals occurred either as clusters of needles or as separate rod-shaped crystals, often together with precipitate (Fig. 54).

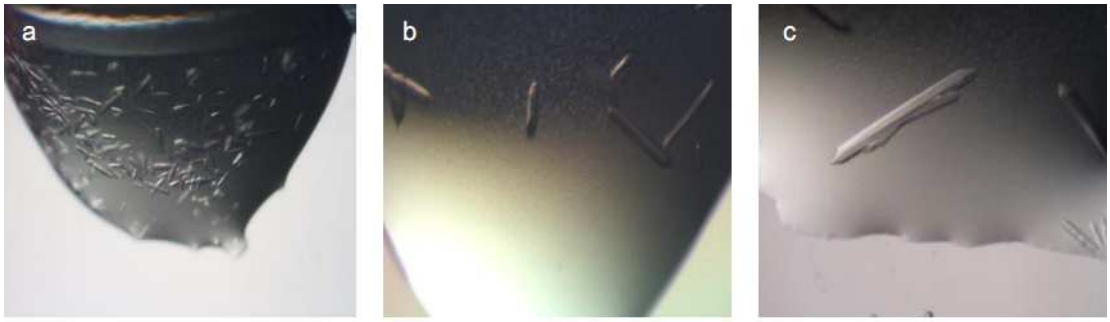


Fig. 54: Initial SpirTE crystals. a: 0.1 M MES pH 6.5, 10 5 b: 0.2M NaCl, 0.1 M Na-phosphate pH 6.2, 10 % (w/v) PEG8000, c: 0.2 M Na-acetate, 0.1 M Na-citrate, pH 5.5, 5 % PEG4000

Optimization of the crystallization conditions (10 % PEG8000, 0.15 M NaCl, 0.2 M Na-phosphate pH 6.2) in 24-well hanging drop plates and seeding resulted in large rod-shaped crystals which diffracted to about 3 Å (Fig. 55).

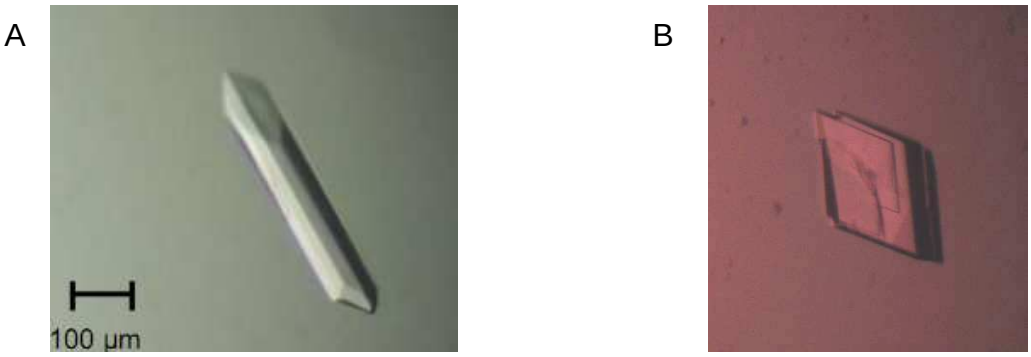


Fig. 55: Optimized SpirTE crystals. A: rod-shaped crystal; B: rhombohedrical crystal. Both from 10% PEG8000, 0.15 M NaCl, 0.2 M Na-phosphate pH 6.2

SeMet-labelled protein was used to obtain crystals for anomalous scattering experiments to obtain phase information. These, however, sometimes had a different morphology than the original crystals. They were of rhombohedrical shape and often displayed macroscopic twinning. Seeding experiments resulted in large, single crystals, one of which diffracted up to 2.5 Å.

4B.2.3 Data collection

A native dataset of a rod-shaped SpirTE crystal was collected at the beamline ID29 at the ESRF (Grenoble), which diffracted up to 2.9 Å. The space group was determined to be C222 with the unit cell constants a=121.3 Å, b=168.9 Å, c=170.5 Å. Molecular replacement was tried with several homologues such as 1KEZ (DESB TE) and 2HFK (PIK TE) as well as with parts of these structures, but no solution could be obtained. Thus, a SAD dataset of a SeMet-labelled crystal with a rhombohedrical shape was measured at the same beamline. This crystal grew in the space group $P2_12_12$ with the unit cell constants a=99.5 Å, b=167.0 Å, c=41.2 Å. Data processing was carried out using XDS (Kabsch 2010) and scaling using XSCALE. The Matthews coefficient (Matthews 1968) was calculated to VM = 2.86 $Å^3$ Da^{-1} indicating two monomers per asymmetric unit corresponding to a solvent content of 57%. Heavy atom positions and initial phase information were obtained by ShelX (Sheldrick 2008) which detected 6 of 10 possible selenium sites. These heavy atom sites were used in Phaser (McCoy *et al.* 2007) for heavy atom site refinement and detecting 3 additional sites. The resulting phases were subjected to density modification with Parrot (Cowtan 2010) and a model was built into the improved density using Buccaneer (Cowtan 2006) which built most of the two protein chains. Manual model building and

refinement were performed with Coot (Emsley *et al.* 2010) and Refmac5 (Murshudov *et al.* 1997). In the end waters were placed by Coot and a final TLS refinement (Painter & Merritt 2006a) step using five TLS groups per monomer as determined by the TLS Motion Determination Server (Painter & Merritt 2006b) was carried out with REFMAC5. The final R values were 21.2 %/ 27.2%, (Rwork / Rfree), respectively.

Table 13: SpirTE data collection and refinement statistics

	SAD
X-ray source	ESRF ID29
Space group	$P2_12_12$
Unit cell dimensions *a*, *b*, *c* (Å)	99.6, 167, 41.3
Wavelength (Å)	0.9793
Resolution (Å)	64-2.45 (2.58-2.45)
Measured reflections	355450 (51838)
Unique reflections	26249 (3790)
I/σ_I	14.8 (3.8)
Completeness (%)	100 (100)
Redundancy	13.5 (13.8)
R_{merge} (%)	9.3 (56.5)
Wilson B factor (Å^2)	67.7
Solvent content (%)	57%
R_{work} (%)	21.2
R_{free} (%)	27.2
Protein atoms	4023
Solvent atoms	86
r.m.s.d. from ideal geometry:	
Bond lengths (Å)	0.0216
Bond angles (°)	1.698
Average B-factor (Å^2):	
Protein (chain A)	22.2
Protein (chain B)	22.9
Solvent	23.3
Ramachandran plot (%):	
Favoured region	93.7
Allowed region	5.6

4B.2.4 SpirTE structure

The asymmetric unit of the SpirTE crystals contains two monomers. These superimpose very well as shown by an rmsd of 0.53 Å for C_α atoms except for the loop connecting β6 and β7. For both chains residues 21 to 283 could be built. The protein consists of two domains: a hydrolase core domain and an α-helical lid or dimerisation domain (Fig. 56). The hydrolase domain exhibits an α/β-hydrolase fold lacking the first β-strand (Ollis *et al.* 1992). Thus, it consists of a central seven-stranded β-sheet, which is all parallel except for the first strand, surrounded by four

α-helices (α3-5 and α8). The lid domain consists of four helices (α1/2 and α6/7). Helices α6 and α7 pack against helices α1 and α2 in a kind of hash key arrangement. These latter two helices mediate an extensive interface between the two monomers in the asymmetric unit with an interface area of 818 Å^2 and a ΔG of -18.5kcal/mol. This is the only significant interface as judged by the PISA server (Krissinel & Henrick 2007) and has been observed in all other PKS TE domain structures so far (Akey *et al.* 2006; Samel *et al.* 2006; Tsai *et al.* 2001). Thus, this is probably the dimer which has been suggested by gel filtration.

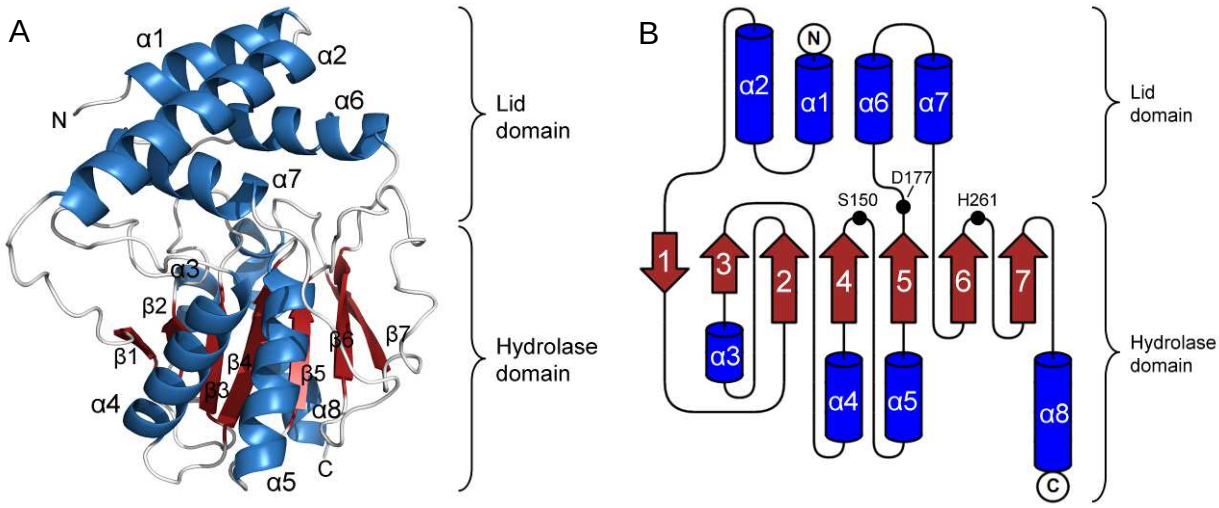


Fig. 56: SpirTE structure. A: Cartoon diagram of a SpirTE monomer colored by secondary structure (helix: blue, strands: red and loop: white); B: Topology diagram; the positions of the residues of the catalytic triad are marked by black circles.

4B.2.5 Active site

In a surface representation a tunnel between the two domains of the protein is visible (Fig. 58A and B) which spans the entire protein. This tunnel is roughly 25 Å long and in the middle about 6 Å in diameter. The active site of the protein is located approximately at the centre of this tunnel. SpirTE has the catalytic triad Ser150, His261 and Asp177. Ser 150 is positioned on a loop after β4 and Asp177 on a loop after β5 while His 261 is found on a loop after β7 (Fig. 56B). The aspartate forms a hydrogen bond to the histidine which thereby becomes more eletronegative (Fig. 57). The histidine can thus abstract the proton from the serine which enables the serine to attack the thioester bond of the substrate. The backbone nitrogen of residue S151

might act as an oxyanion hole, which donates a hydrogen bond to and thereby stabilises the negatively charged oxygen at the tetrahedral intermediate.

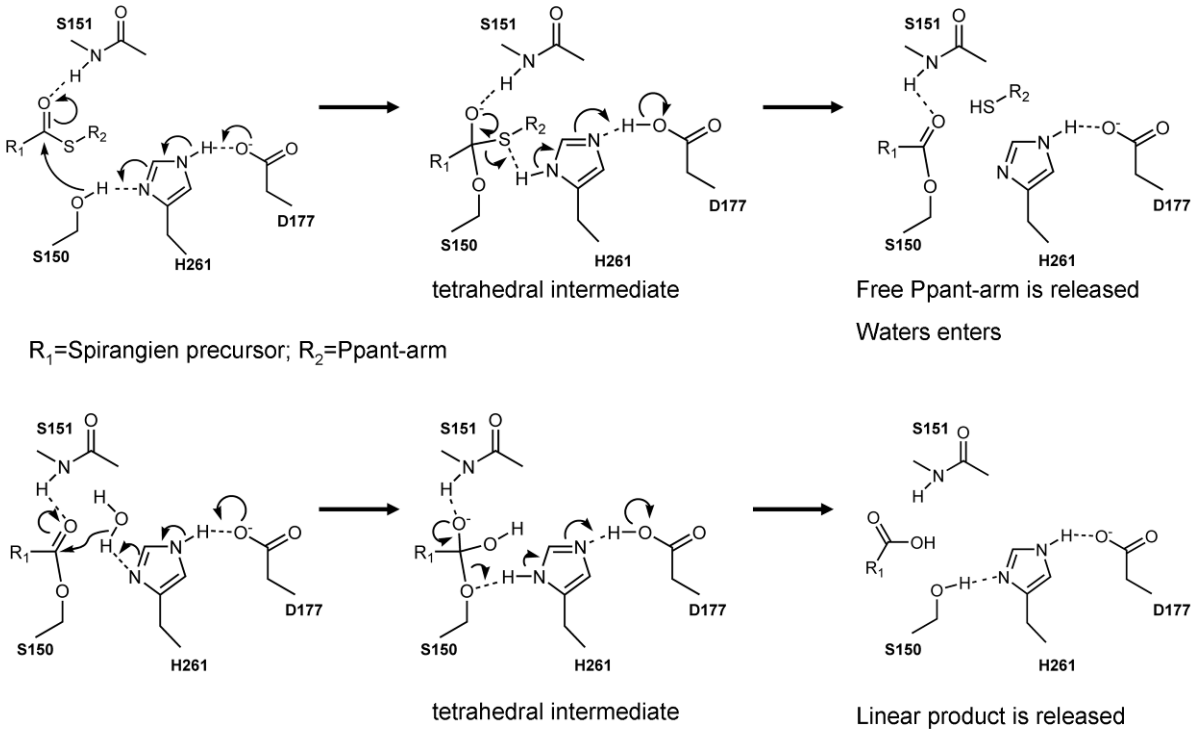


Fig. 57: Putative reaction mechanism of SpirTE. The catalytic triade S150, H261, D177 catalyses the claevage of the thioester bond of the spirangien precursor, while the backbone amide of S151 acts as an oxyanion hole to stabilise the tetrahedral intermediate state.

The tunnel opens at one end in spatial proximity (about 16 Å) to the N-terminus of the protein (Fig. 58A). In the full-length protein (SpiJ) the ACP domain would be located here. Thus, the ACP domain can bind to the entrance of the tunnel and insert the PPant-bound spirangien-precursor into the tunnel. The thioester is then cleaved in the active site and the free spirangien-precursor can leave the protein via the exit tunnel, whereas the empty PPant-arm retracts from the entrance and is ready to take up a new substrate.

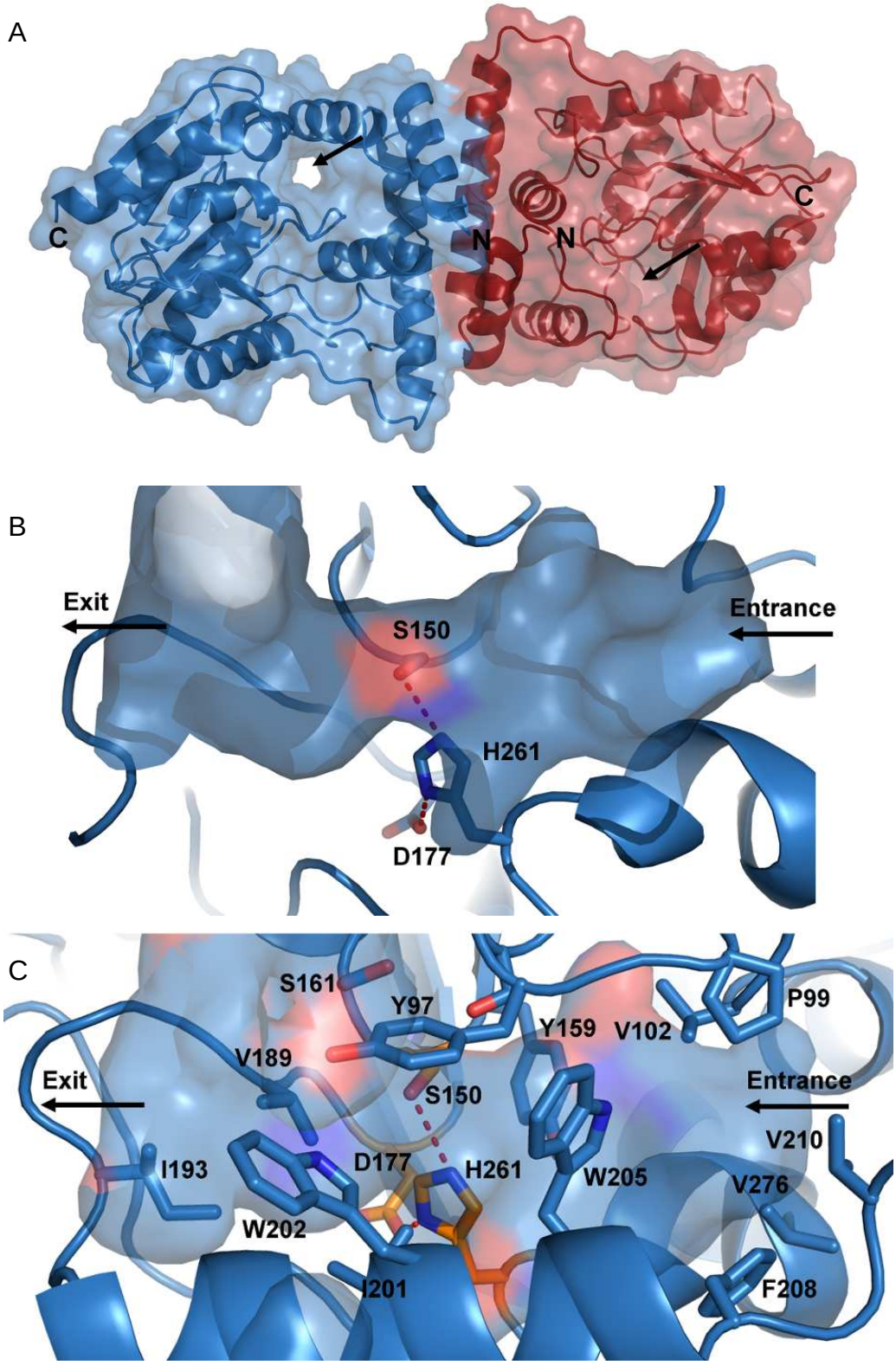


Fig. 58: SpirTE substrate tunnel and active site. A: SpirTE dimer as cartoon and surface showing the tunnel between the lid and the hydrolase domain (black arrows). B: Surface view of the substrate tunnel with the catalytic triade shown as sticks. C: The substrate tunnel is lined with hydrophobic residues (shown as sticks).

The substrate tunnel is lined with mostly hydrophobic residues such as I201, W202, W205, F208, V102, V210, V189 and V276 and only few hydrophilic patches such as those created by Y97 and W202 (Fig. 58C). The first half of the PPant-arm which will be inserted into the tunnel is highly hydrophobic. The same is true for the spirangien-precursor which is the substrate for SpirTE, especially the region that would be bound to the protein. Thus, once the first part of the long and flexible molecule has entered the tunnel, the rest will be pulled into the protein by the force of the hydrophobic interactions.

5 Discussion

5A Virulence regulation in pathogenic Yersiniae

5A.1 The transcription factor RovA

5A.1.1 RovA is a key player in virulence regulation

The transcription factor RovA plays a major role in the pathogenesis of all pathogenic *Yersinia* species (Heroven & Dersch 2010). In enteropathogenic *Yersinia* it is the main regulator in a complex regulatory network controlling the expression of the adhesin invasin which is required for rapid uptake of the bacteria through the host epithelium. RovA has been shown to act as a protein thermometer (Herbst *et al.* 2009). Thus, it exhibits strongly reduced DNA binding affinity at 37℃ compared to 25℃ and is degraded proteolytically *in vivo*. To gain additional insight into the complex regulation of RovA in this work it was further characterised using biophysical methods and the structure was solved in complex with DNA and in complex with the potential inducer molecule salicylate.

5A.1.2 Small aromatic compounds act as inducers for RovA

For several MarR-type transcription factors small inducer molecules have been described that bind to the proteins and thereby influence their DNA binding capability (Wilkinson & Grove 2006). These include small aromatic compounds such as salicylate (Alekshun & Levy 1999), uric acid (Wilkinson & Grove 2004) and even chlorinated compounds (Providenti & Wyndham 2001). For RovA no inducer molecule has been described so far. Therefore, several small compounds were tested for their affinity to RovA by Thermofluor assay. These assays showed that RovA binds to small aromatic acids such as salicylate and benzoate with millimolar affinity. Additional hydroxyl groups as in 2,3-dihydroxybenzoate or 4-hydroxybenzoate bind with similar affinity, whereas amino groups as in anthranilate or 4-aminobenzoate resulted in a reduced affinity and phenylalanine did not bind at all. Additionally, uric acid was tested and also found not to bind to RovA. Surprisingly, cinnamate and 3-chlorocinnamate were found to bind to RovA with higher affinity

than the other compounds. Often, the compounds that bind to MarR-type proteins are related to the genes they regulate such as chlorobenzoate, which is an inducer of CbaR which regulates genes involved in the degradation of chlorobenzoate (Providenti & Wyndham 2001). Unfortunately, for RovA no such relationship is evident. Therefore, it is still unclear which compounds act as inducers for RovA *in vivo*. To understand the effect of inducer binding to the function of RovA, gel shift assays in the presence of salicylate were performed. These showed that addition of salicylate to RovA reduces its DNA binding affinity and thereby acts as an inhibitor of RovA. Similar results have been obtained for other MarR-type proteins such as MarR from *E. coli* (Martin & Rosner 1995), ST1710 from *Sulfolobus tokodaii* (Kumarevel *et al.* 2009) and MTH313 from *Methanobacterium thermoautotrophicum* (Saridakis *et al.* 2008). CD spectroscopy was used to analyse the impact of salicylate binding on the conformational flexibility of RovA. RovA shows a slow, reversible decrease of α-helix content during an increase in temperature from 20℃ to 37℃. Upon addition of salicylate, however, this change is strongly reduced. This is probably because RovA is stabilised and locked in a certain conformation when bound to salicylate and thus less prone to conformational flexibility. A stabilisation RovA was also observed *in vivo* by the addition of salicylate to *Y. pseudotuberculosis* cells. This indicates that a small aromatic compound, similar to salicylate, might play a role in the regulation of RovA production.

In vivo, a stabilisation of RovA has been observed in bacteria cultures that have reached the stationary phase (Herbst *et al.* 2009). Here, even at 37℃ RovA is degraded only very slowly. Experiments have been conducted to show that this stabilisation is conferred by the medium in which the cells were grown (Herbst, unpublished results). To this end, Y. pseudotuberculosis cells were grown to stationary phase and the removed from the medium by centrifugation. The cleared medium was then used to grow bacteria in the exponential phase. These bacteria now also showed a stabilised RovA as if they had been grown in stationary phase. These results suggest that there might be a small molecule that is secreted into the medium by bacteria growing in stationary phase that stabilises RovA. Whether this molecule directly binds to RovA and the chemical nature of this molecule are so far unknown and will be subject of further research. It is also possible that this or a similar molecule is also produced by E. coli and was visible as the undescribed density in the RovA/DNA structures.

5A.1.3 The RovA mutant P98S is proteolytically stable

While the sequence of RovA is identical in *Y. pestis* and *Y. pseudotuberculosis*, the *Y. enterocolitica* RovA is slightly different. One of the mutations (P98S) has been shown to lead to a stabilisation *in vivo* and it was proposed that this mutation might influence the thermosensing ability of RovA. Thus, the DNA binding, salicylate binding and thermal stability have been tested *in vitro*. Together, these tests showed that RovA P98S behaves exactly like the wild type *in vitro*. Therefore, it seems probable that the difference in stability observed *in vivo* is due to a reduced proteolytic degradation. Possibly, this mutation hampers the recognition of RovA by the protease.

5A.1.4 RovA A10F, a stable mutant *in vitro*

As the binding of small molecules such as salicylate to RovA has been shown to have a strong impact on the behavior of RovA *in vitro* and *in vivo*, mutants of RovA were generated in order to block the ligand binding cavity in RovA. Towards this end, two small amino acids (G6 and A10), which line the cavity, were selected because their *in silico* mutagenesis to phenylalanine promised to block ligand binding in the cavity (Fig. 59A and B). Both mutants were produced and tested for their stability and salicylate binding capability by ThermoFluor and CD measurements. While the G6F mutant behaved like the wild type, the A10F mutant showed a significantly higher melting temperature than the wildtype. The melting temperature of RovA A10F in the absence of salicylate is comparable to the melting temperature of the wild type in the presence of 20mM salicylate. RovA A10F still bound salicylate although with a much lower affinity than the wild type. Additionally, the CD spectra of RovA A10F at 20°C and 37°C indicate a strongly deccreased conformatio nal flexibility of this mutant. Therefore, it seems that the phenylalanine sidechain occupies the binding pocket where salicylate can bind. This would stabilise the protein and prevent salicylate binding. The residual salicylate binding might be explained by incomplete blockage of the ligand binding pocket or the presence of another weaker binding site such as visible in the structure of RovA with salicylate. To examine the impact of the blockage of the ligand pocket on the DNA binding capability, further experiments have to be carried out such as gel shift assays and *in vivo* stability assays. The latter might give important insights into the potential role of small molecules in the regulation of RovA.

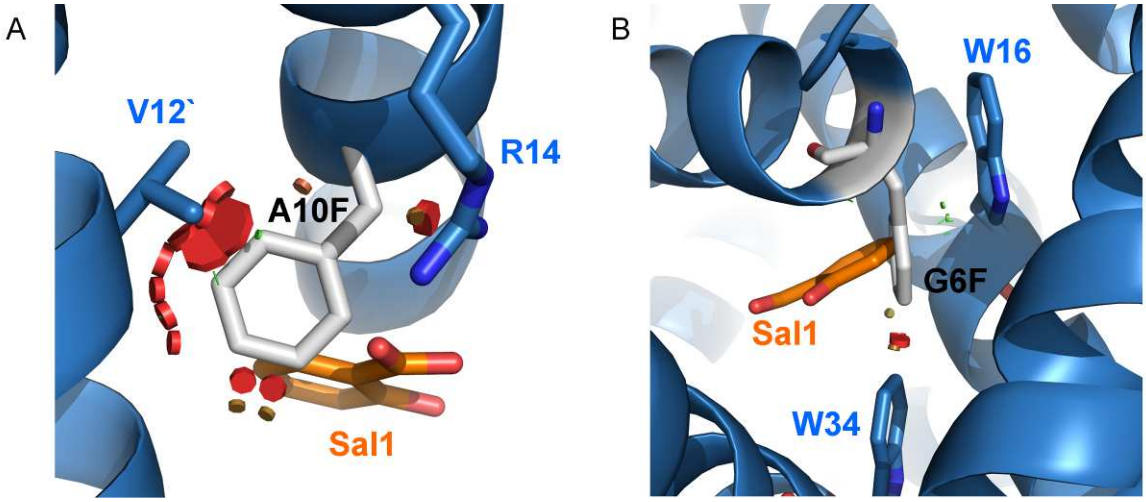


Fig. 59: *In silico* model of mutagenesis of RovA (A: A10F, B: G6F). RovA is shown in blue, salicylate inorange, the mutated residue in white. Steric clashes with neighboring residues are marked by disks.

5A.1.4 RovA recognizes DNA in a similar fashion as OhrR

In this work the first structure of the global virulence regulator RovA from *Y. pseudotuberculosis* has been solved in complex with DNA fragments from the *rovA* and *inv* promoter region. Two structures of MarR-type proteins in complex with DNA have been published so far, OhrR from *B. subtilis* (Hong *et al.* 2005) and ST1710 from *Sulfolobus tokodaii* (Kumarevel *et al.* 2009). While the overall DNA binding mechanism is similar between RovA and OhrR, with the recognition helix of the HTH motif inserted into the major groove of the DNA and the wing contacting the minor grove (Fig. 60A), the DNA binding observed for ST1710 seems to be quite different (Fig. 60B). In the structure of ST1710, the recognition helix is not inserted into the major groove and thus makes only few base-specific contacts with the DNA, whereas most contacts are mediated by the wing. Yet, even though the structures of RovA and OhrR in complex with DNA are similar, the rmsd of 3.8 Å for common Cα atoms is relatively high, mostly due to differences in the long helices of the dimerisation interface (Fig. 60A).

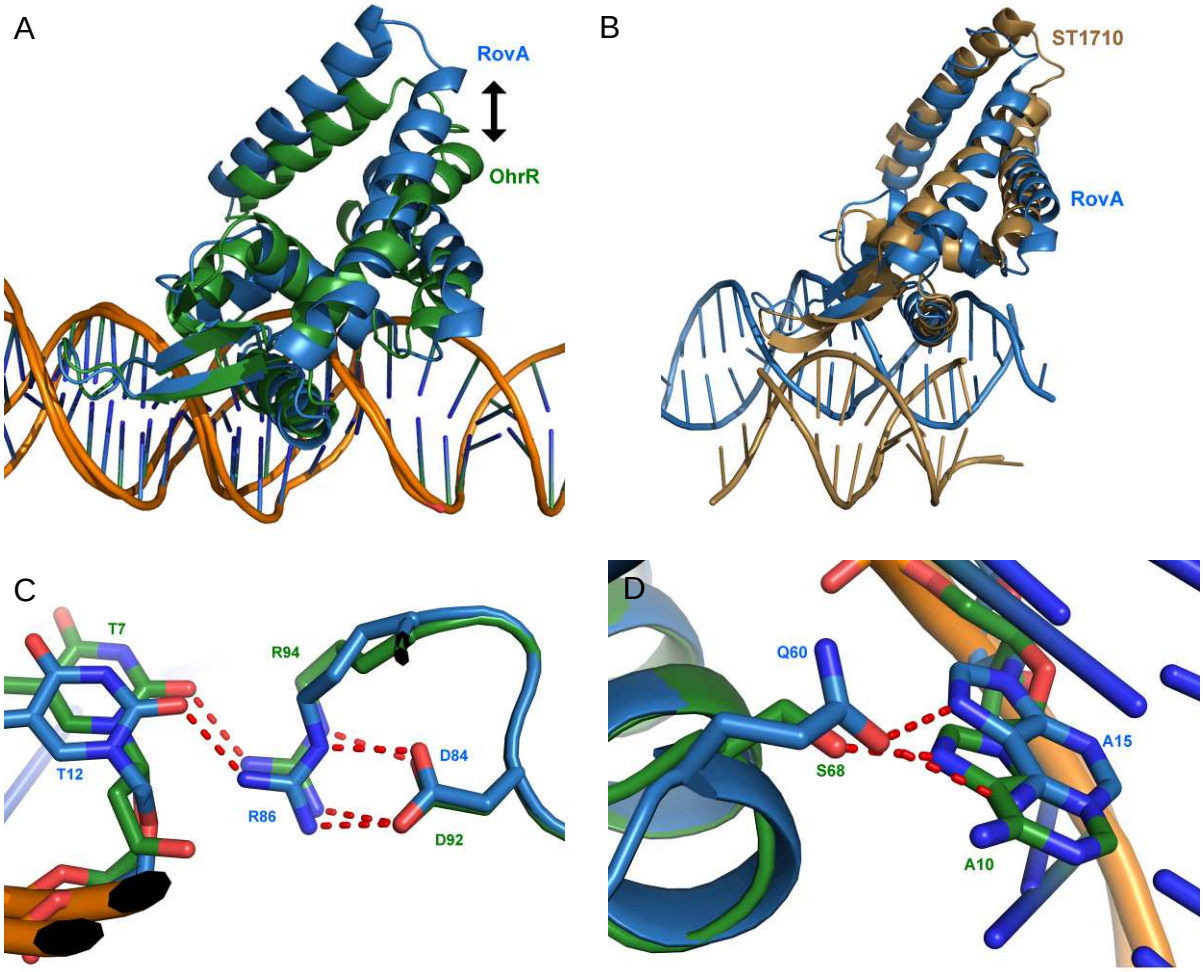


Fig. 60: Structural alignment of RovA (blue) and OhrR (green) or ST1710 (brown) in complex with DNA. A: Comparison of one monomer of RovA and OhrR aligned by their DBDs showing differences in the long helices of the dimerisation domain (indicated by a black arrow). B: Comparison of one monomer of RovA and ST1710 showing the superficial DNA binding of ST1710 compared to RovA. C: A similar DNA binding motif is employed by RovA and OhrR involving an arginine and an aspartate residue. D: Different amino acids are used to establish contacts to the DNA at the same positions.

While the overall DNA binding by RovA and OhrR is similar, the amino acids that are involved in base recognition are different, reflecting their individual DNA specificity (Fig. 61). OhrR interacts with the bases of its DNA specifically with six amino acids, namely D67, S68, G69 and T70 from helix 4 and R88 and R94 from the wing. Only R94, which contacts the DNA in the minor groove, is also present in RovA (R86) (Fig. 60C). It is also conserved in many other MarR-type proteins and may therefore not necessarily be used to gain specificity but rather increase the general DNA binding affinity (Hong *et al.* 2005). Interestingly, even though the other amino acids that make specific contact to the DNA are different, the same positions are used to establish

DNA contacts. For example S68 in OhrR recognizes an adenine residue by a hydrogen bond while RovA also recognizes an adenine residue at the same position, but with a different amino acid (Q60) which is located at the same position in therecognition helix (Fig. 60D). Therefore, it seems that the shared fold of the DBD allows certain positions to establish contact to the DNA. These positions probably mainly define the DNA binding specificity of the protein.

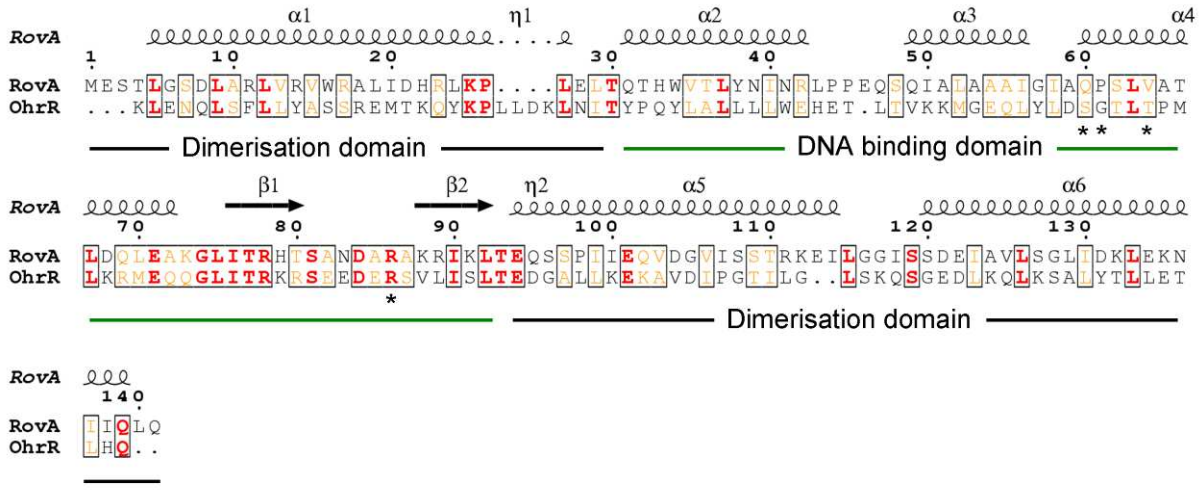


Fig. 61: Structure based sequence alignment of RovA with OhrR from *Bacillus subtilis*. The sequences of RovA from *Y. pseudotuberculosis* and OhrR were aligned with ClustalW (Larkin *et al.* 2007) and displayed with ESPript2.2 (Gouet *et al.* 1999). The dimerisation domain is shown by a black line, the DNA binding domain by a green line. Residues found to recognize DNA bases specifically are denoted by asterisks.

5A.1.5 RovA is a global transcription regulator

Overall, RovA does have less amino acids that specifically interact with the DNA than OhrR. This reflects the role of RovA as a general transcription factor, responsible for the regulation of several genes in the genome (Ellison *et al.* 2004). This is in contrast to many other MarR-type regulators such as OhrR which only binds to one defined binding site, which is nearly perfectly palindromic (Fuangthong *et al.* 2001) and MarR which binds to two nearly palindromic sequences (Martin & Rosner 1995). For RovA, however, no defined palindromic sequence could be found. Instead, RovA was found to prefer AT-rich sequences (Heroven *et al.* 2004). Furthermore, in this work the structure of RovA was solved in complex with two different DNA fragments which have only 11 out of 21 identical bases. This further underlines the variable nature of the DNA recognition of RovA. Additionally, it was found that OhrR and MarR bind to

their target DNA with high affinity (K_D=5nM and 1nM, for OhrR and MarR, respectively) (Fuangthong & Helmann 2002; Martin & Rosner 1995) whilst , for example, the affinity of RovA with the *inv* promoter region is much lower (K_d=32nM and 45nM for binding site I and II, respectively) (Herbst *et al.* 2009). Therefore, it can be concluded that RovA interacts with the DNA specifically only at few positions. This explains the promiscuity of the DNA binding which is required for its role as a global transcription factor.

5A.1.6 RovA relieves silencing by H-NS

In contrast to OhrR which acts as a repressor by binding in the promoter region of its target gene, RovA acts as an activator of *inv* as well as its own gene. This activation is achieved by displacement of the repressing nucleoid-associated protein H-NS (Ellison & Miller 2006; Heroven *et al.* 2004; Tran *et al.* 2005). RovA and H-NS compete for binding to overlapping sequences in the promoter regions of *rovA* and *inv*. H-NS binding to DNA has been shown to lead to bending, compaction and also to bridging between distant DNA strands leading to repression of transcription of adjacent genes (Dame *et al.* 2005; Noom *et al.* 2007). An additional factor that is needed for repression of *rovA* transcription is the LysR-type transcription regulator RovM which is also suspected to induce bending of the *rovA* promoter region (Heroven & Dersch 2006). It has been suggested that the binding of H-NS and RovM to DNA might lead to the formation of a stable inhibitory complex which loops the DNA and thereby prevents binding of the RNA polymerase (Heroven & Dersch 2006). On the other hand, the RovA DNA complex shows the DNA in a linear conformation. This suggests that the inhibitory H-NS/RovM complex can be displaced by elevated concentrations of RovA. This would alleviate the bending of the DNA by the enforced linear conformation, where RovA has bound and thereby allows RNA polymerase binding and gene transcription.

5A.1.7 Comparison of RovA and SlyA DNA binding sequences

The structure of RovA in complex with DNA is overall similar to the yet unpublished structure of the close homologue SlyA from *Salmonella typhimurium* (pdb code: 3DEU, 75% sequence identity) in complex with salicylate with an rmsd of 1.65 Å for common C_α atoms. Similar to RovA, SlyA has been shown to be a global

transcription factor regulating the expression of a large number of genes involved in virulence and resistance to oxidative stress (Buchmeier *et al.* 1997). Nearly all amino acids which are different between these two proteins are on the surface, mainly on the two C-terminal helices (Fig. 63). None of them is directly involved in DNA binding. We thus propose that SlyA and RovA should exhibit the same DNA binding specificity. However, the DNA consensus binding sequence of SlyA that has been published is quite different from the sequences in the RovA structures (Stapleton *et al.* 2002). Most notably, at positions 1 and 12, in which an adenine residue is contacted by Q60, the SlyA consensus sequence has a thymidine (Fig. 62), which would not be recognized by the protein as in one monomer of the RovA/rovA1A structure. At positions 4, 5, 8 and 9 a strong preference for G or C has been determined for SlyA, whereas these bases occur only rarely in RovA binding sequences. These differences are surprising as the DBDs of RovA and SlyA are nearly identical. Therefore, additional experiments have to be carried out to investigate these observed discrepancies.

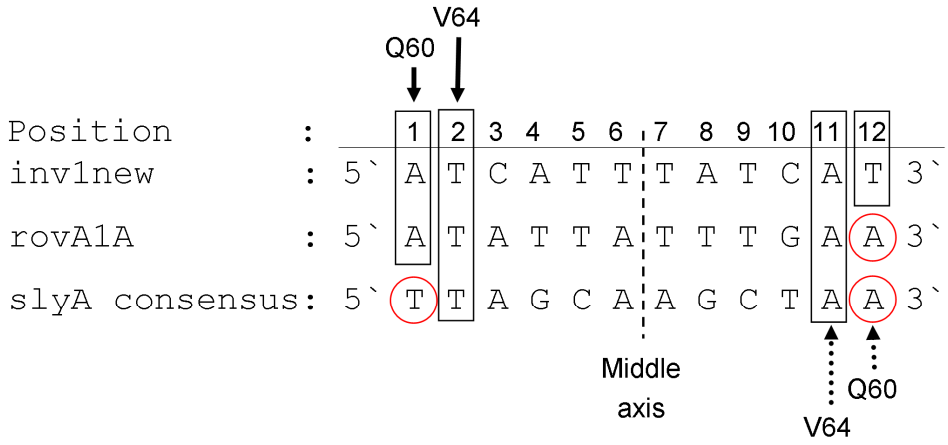


Fig. 62: Comparison of the DNA sequences from the RovA/DNA structures compared with the consensus sequence for SlyA (Stapleton *et al.* 2002). The positions that are specifically recognized by amino acids in the RovA structures are marked by boxes. Deviations at these positions are marked by red circles.

5A.1.8 Mutations in RovA and SlyA defective in DNA binding

Several mutants of RovA and SlyA defective in DNA binding have been described, which can now be explained by the presented structure. For RovA single mutations in residues 44, 51, 60, 61 and 71 (Tran *et al.* 2005) and in SlyA mutations of amino

acids 52, 53 60, 63, 64, 65, 67, 70, 71, 74, 75, 77, 78, 79, 80, 85, 86, 88 and 90 were found to strongly reduce DNA binding affinity (Haider *et al.* 2008). All of these SlyA residues are conserved in RovA, except S77 which is a threonine in RovA and Q79 which is a histidine in RovA (Fig. 63).
P44 is located in a loop between helices α2 and α3 and is probably required for proper positioning of these helices. The mutation Q51R shows only a mild effect on DNA binding. It is not clear why it would impact DNA binding, as the residue is disordered in the DNA complex structure. P61 is positioned at the end of helix α4 and recognizes the DNA via hydrophobic interactions. A mutation of this residue to leucin would lead to strong steric interactions with the DNA and was thus shown to completely abolish DNA binding.

Amino acids A53, L52, L63, L67, L70 and I90 form the hydrophobic core of the wHTH domain and are thus probably indispensable for its stability. G74 and L75 are located on the loop between α4 and the first β-sheet of the wing (β1) and are responsible for correct positioning of the wHTH domain by hydrophobic interactions as well as hydrogen bonds with helix α6 from the dimerization domain. V64 could be seen to be responsible for hydrophobic interactions with a thymidine residue at position 5 from the centre of pseudosymmetry. The mutation V64I probably leads to a steric hindrance with the DNA und thus to reduced affinity. The important role of Q60 and R86 in recognizing bases has been described above and is reflected by a complete loss of binding affinity for the respective alanine mutants in gel shift experiments. R65, on the other hand, which also seems to be crucial for DNA binding *in vitro*, is not visible in our structures in complex with DNA due to flexibility. It is well positioned to allow binding to bases in the major groove of the DNA and it may be necessary for binding to a base which is not present in this position in our structure. R78 and K88 are directly involved in hydrogen bonds with the phosphate backbone, while E71 interacts with to R78 and thus keeps it in place. Q79 (H79 in RovA) might be important for keeping the wing in place as it forms a hydrogen bond with R89 from the other β-strand of the wing. The role of S77 (T77 in RovA) and T80, on the other hand, is not visible from the structure.

In conclusion, the role of most of the mutations that have been identified to hamper RovA or SlyA DNA binding can be explained by the structure presented in this work.

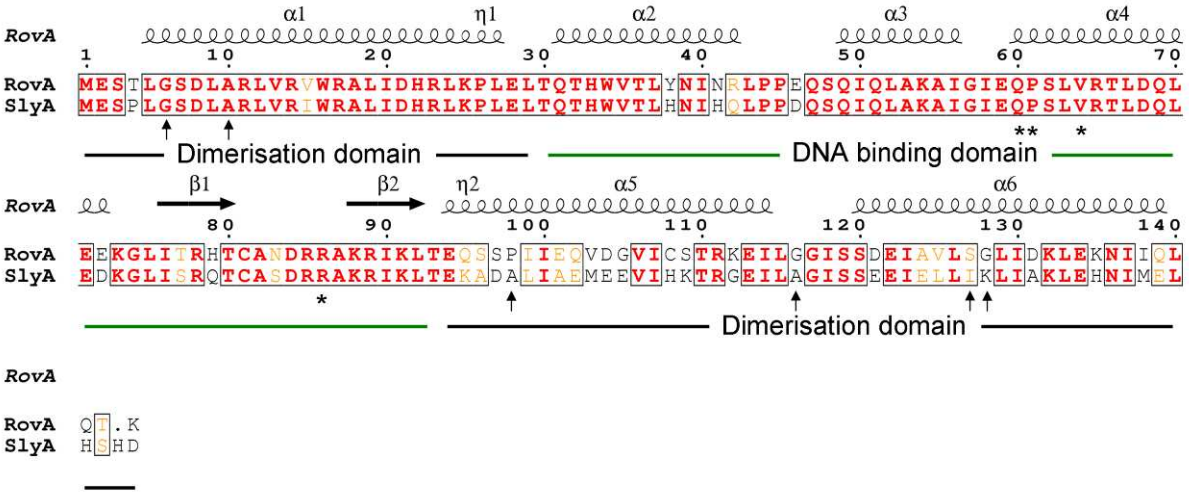


Fig. 63: Structure based sequence alignment of RovA with SlyA from *S. typhimurium*. The sequences of RovA from *Y. pseudotuberculosis* and SlyA were aligned with ClustalW (Larkin *et al.* 2007) and displayed with ESPript2.2 (Gouet *et al.* 1999). The dimerisation domain is shown with a black line, the DNA binding domain by a green line. Residues found to recognize DNA bases specifically are denoted by asterisks and residues which were mutated in this work are marked by arrows.

5A.1.9 The RovA-apo structure shows conformational flexibility

The structure of RovA-apo could only be solved by molecular replacement using the structure of RovA in complex with DNA. The molecular replacement found five monomers in the asymmetric unit and another monomer was detected and placed manually during refinement. However, this monomer showed only weak or no electron density in those parts of the protein which are not directly involved in dimer formation, namely the DBD. Comparison of the RovA monomers showed that the conformation of RovA is probably quite flexible without stabilising binding partners such as salicylate or DNA. Interesting is also the superposition of the structure of RovA-apo and the DNA bound form (Fig. 64A). The rmsd for all common C_α atoms between the RovA in complex with the inv1 DNA and RovA apo is 1.1 Å / 2.4 Å / 2.3Å for RovA-apo dimer AB/CD and EF, respectively. This demonstrates that the dimer AB has a conformation which is most similar to the DNA bound form, while dimers CD and EF are more different. The superposition of one chain of each dimer shows that the DBD of the other chain has moved in the apo structure relative to the DNA-bound form. This conformational change places the DBD closer towards the DNA and reduces the gap between the two DBDs in the dimer. The movement is more

pronounced for RovA-apo dimers CD and EF with a distance of 6.4 Å and 6.8 Å, respectively, between the C_α atoms of Q60, an important residue in the recognition helix, compared to DNA-bound RovA. This distance is only 4.3 Å for RovA-apo dimer AB. Similar conformational flexibility has been observed also for other MarR-type transcription factors such as MexR (Lim *et al.* 2002). As the conformation of all RovA-apo dimers would lead to a steric clash with the DNA, it can be concluded that RovA without ligands can probably perform a twisting motion (similar to the vibrational mode in a water molecule) and only the open form is able to bind to DNA (Fig. 64A). Interestingly, a different, scissoring conformational change between the apo form and the DNA bound form has been described for OhrR (Fig. 64B) (Hong *et al.* 2005). Here, the apo form assumes a more open conformation, moving the DBD away from the DNA, rather than closer to the DNA. Therefore it seems that different conformations are used by proteins from the same family.

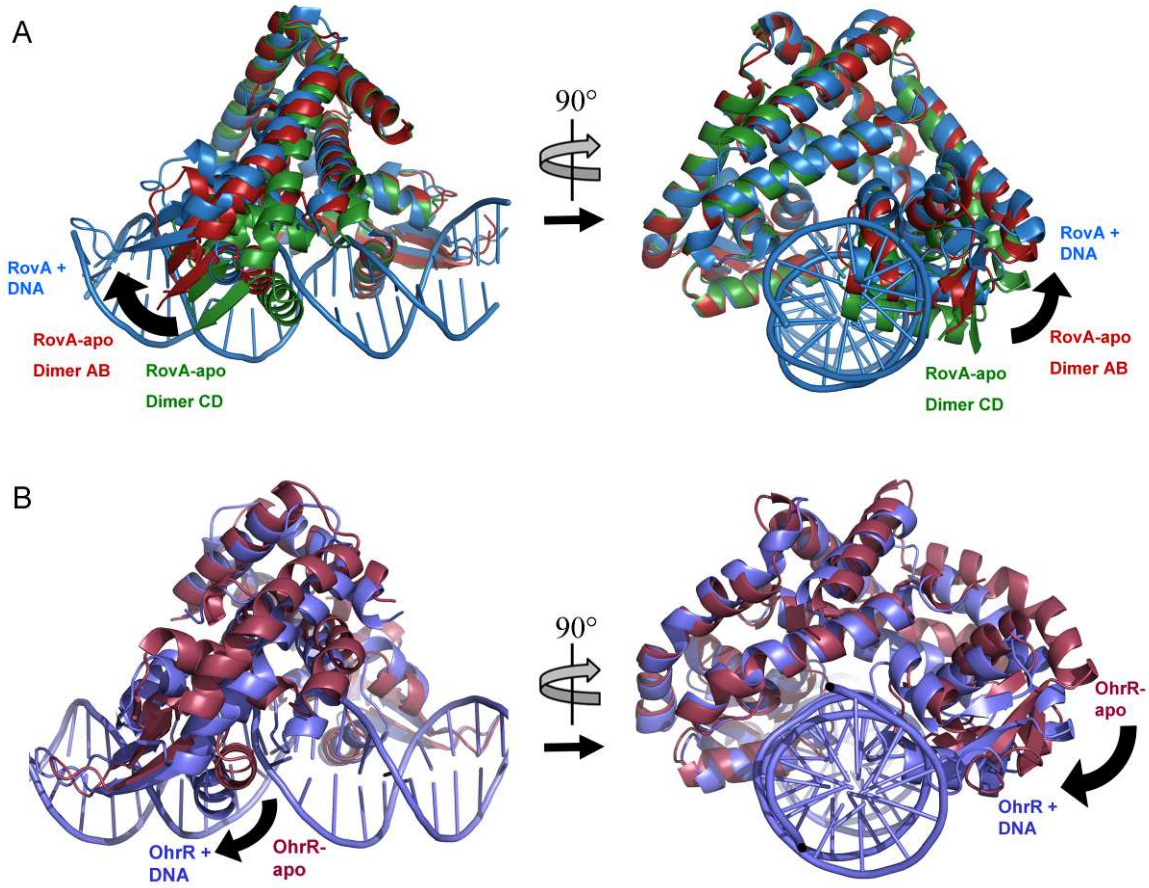


Fig. 64: Conformatonal movements upon DNA binding of MarR-type proteins (indicated by black arrows). A: Superposition of one chain of RovA/DNA (blue) and RovA apo dimers AB (red) and CD (green). B: Superposition of one chain of OhrR/DNA (blue) and OhrR-apo (red).

5A.1.10 Structure of RovA in complex with salicylate

The structure of RovA in complex with salicylate was solved by using the unpublished structure of SlyA from *Salmonella typhimurium* as a model for molecular replacement. Comparison of the structures of RovA/sal with RovA/DNA shows that both adopt a similar fold. However, a conformational change can be observed which results in as rmsd for all common C_α atoms of 2.3 Å (Fig. 65A). RovA/sal exhibits a closed conformation where the DBDs are closer together than in the DNA bound structure. In order to bind DNA the protein would have to open up in a scissoring motion. A similar motion was detected for OhrR between apo and DNA bound form (Fig. 64B), only that the apo form was in a more open conformation than the DNA bound form.

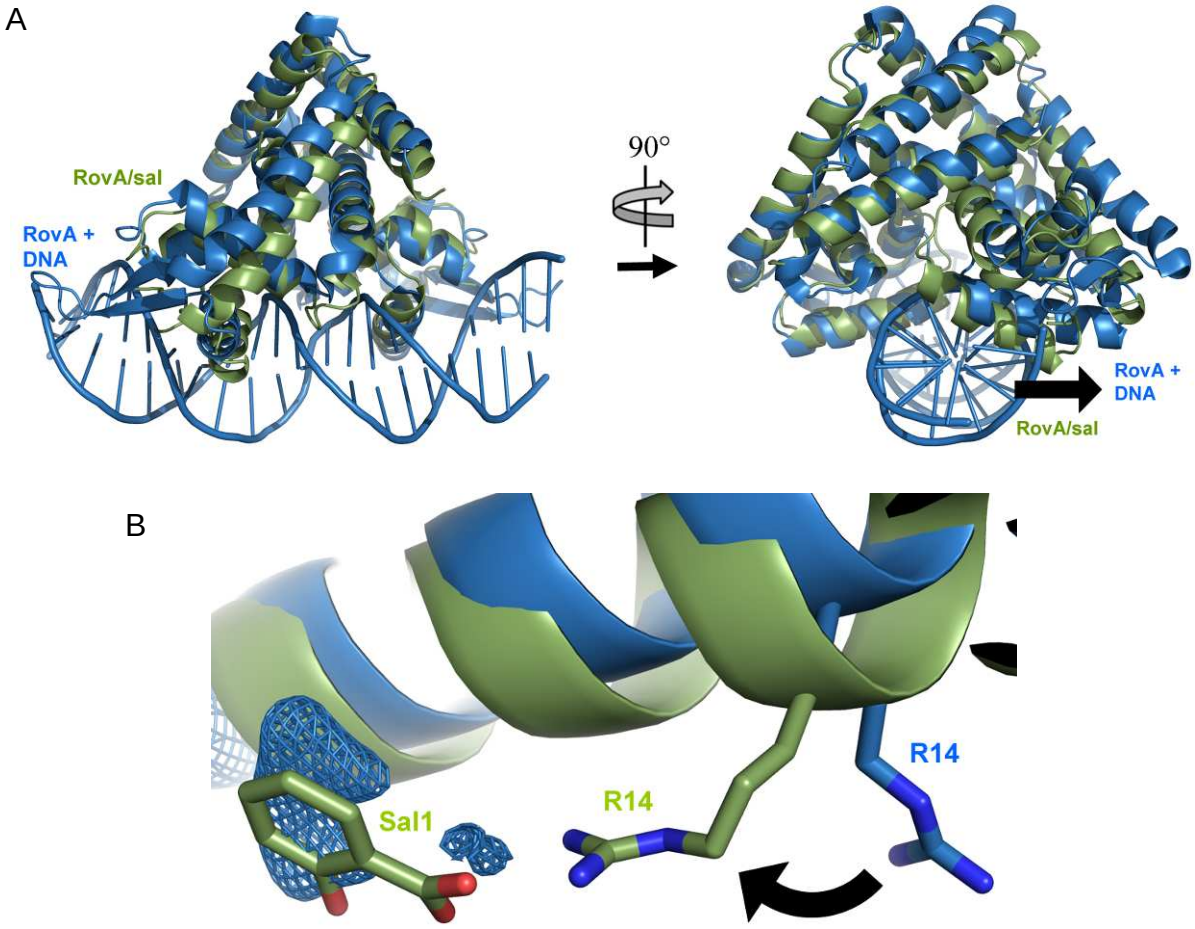


Fig. 65: Conformational differences between RovA bound to DNA (blue) and salicylate (green). A:.Side- and frontview of overall structures aligned by one monomer showing the scissoring motion that has to occur for RovA/sal to allow DNA binding (indicated by black arrow). B: Detailed view of the ligand binding pocket. The F_O-F_C map of RovA/DNA is shown as blue mesh at 4 sigma.

Two salicylate molecules are bound to every RovA monomer. One salicylate molecule was found at the surface of the DBD, the other in a mostly hydrophobic pocket at the dimerisation interface. It is interesting to note that the undescribed density which was detected in the structures of RovA in complex with DNA was found in the same pocket (Fig. 65B). However, it was not possible to explain the density with any of the substances used in the crystallization condition. RovA has been shown in this study to bind to a variety of small hydrophobic molecules and it is therefore possible that such a molecule from the expression host has been carried into the crystallisation. The superposition of the hydrophobic pocket of RovA/sal and RovA/DNA also reveals that R14 which is oriented directly towards the carboxy-group of the salicylate molecule in the RovA/sal structure points into a different direction in the RovA/DNA structure (Fig. 65B). Obviously, the sidechain of R14 rotates towards the pocket to compensate the negative charge of the carboxy-group of the salicylate.

The hydrophobic pocket seems to be involved in ligand binding in many MarR-type proteins. Several proteins from this class such as ST1710, MTH313 or SlyA (Kumarevel *et al.* 2009; Saridakis *et al.* 2008) have been crystallized with salicylate and salicylate molecules were bound in this pocket in most of them (Fig. 66). Additionally, superposition of RovA with OhrR, which is regulated by the oxidation of the reactive cysteine 15, shows that this cysteine is located in direct vicinity to the salicylate molecule of RovA. It is therefore probable that conformational changes at this position either by binding of a ligand such as salicylate (as in RovA) or amino acid modification (as in OhrR) represent a conserved mechanism to influence the DNA binding affinity of MarR-type proteins.

The second salicylate binding site, however, might be an artefact of the unnaturally high salicylate conditions. The salicylate is held in place by only weak undirected hydrophobic interactions and this binding site is also visible in the structure of SlyA (Fig. 66). The structure of SlyA, which shares a high sequence identity with RovA, is very similar to the RovA structure with an rmsd for common C_α atoms of 1.3 Å. It has salicylate molecules bound in the same positions as RovA, however it also has another molecule bound to the surface of the dimer interface. Interestingly, so far, only for MarR have salicylate molecules bound to the DBD been described. Yet,

these are bound in positions distinct from the binding site in RovA. It is therefore questionable whether the observed salicylate binding to the surface of the DBD is physiologically relevant.

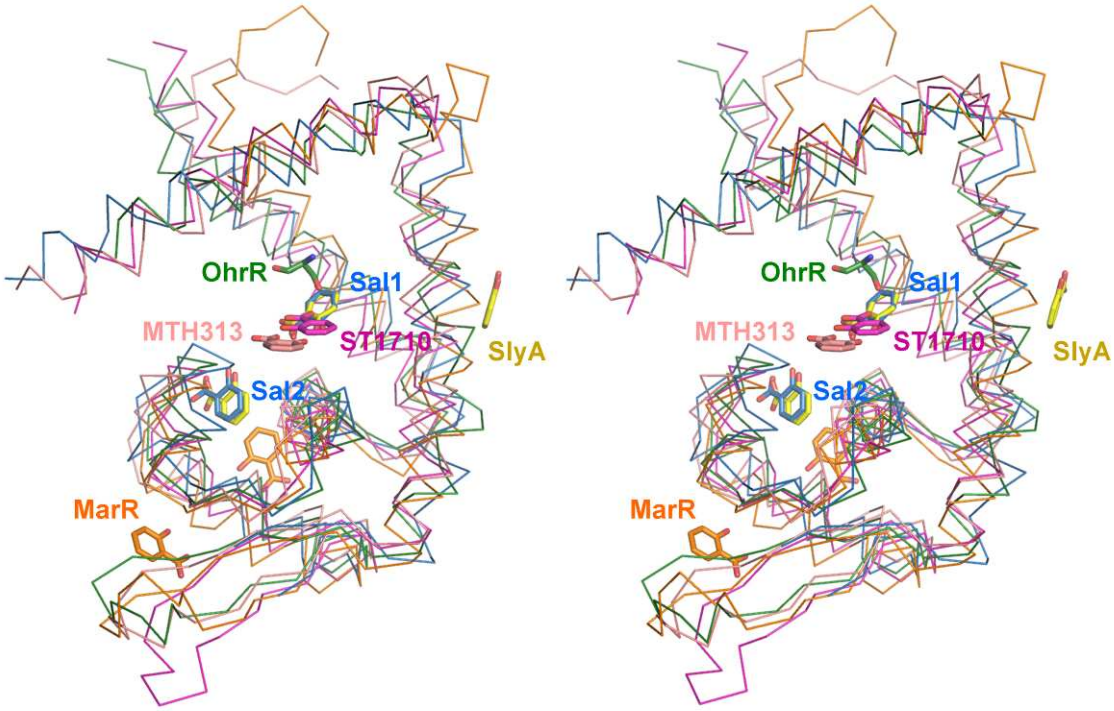


Fig. 66: Comparison of RovA in complex with salicylate with other related proteins. Stereo representation of superposition of one monomer of RovA/sal (blue, Y. pseudotuberculosis) with SlyA (yellow, 3DEU, S. typhimurium), MarR (orange, 1JGS, E. coli), ST1710 (purple, 3GF2, Sulfulobus tokodaii), MTH313 (rose, 3BPX, Methanobacterium thermoautotrophicum) and OhrR (green, 1Z91, B. subtilis). The protein is shown as ribbon, salicylate molcules and C15 of OhrR are shown as sticks.

5A.2 The transcription regulator RovM

5A.2.1 RovM is a repressor of RovA

In addition to the main regulator RovA, there are several other proteins and also RNAs involved in the complex regulatory network regulating *inv* expression. One of these is the LTTR RovM that acts together with H-NS to repress rovA expression (Heroven & Dersch 2006). It links the carbon storage regulator system (CSR), which is thought to be involved in sensing the availability of nutrients, to the RovA system and also influences flagellar motility (Heroven *et al.* 2008). To gain deeper insights into how this protein functions we investigated the structural details of RovM. We were able to produce full-length RovM as well as the effector binding domain of RovM (RovM-EBD), determine their oligomeric state and solve the structure of RovM-EBD.

5A.2.2 Comparison with other LTTR structures

The structure of RovM-EBD (Quade *et al.* 2011) was superimposed with the structure of the EBD of CynR from *E. coli* (Knapp & Hu 2009), with which it shares the highest sequence identity (21%) of all known LTTR EBD structures (Fig. 67A), with an r.m.s.d. of 2.15 Å for all common C_α atoms. Minor structural deviations are only found for exposed loop regions. CynR has been shown to exist as a dimer in solution. Thus, the structure of the CynR dimer was aligned with the BB`- dimer of RovM, which indicates that the BB`-interface is indeed the physiological interface, which is conserved among all LTTR EBDs crystallized to date (Fig. 67B).

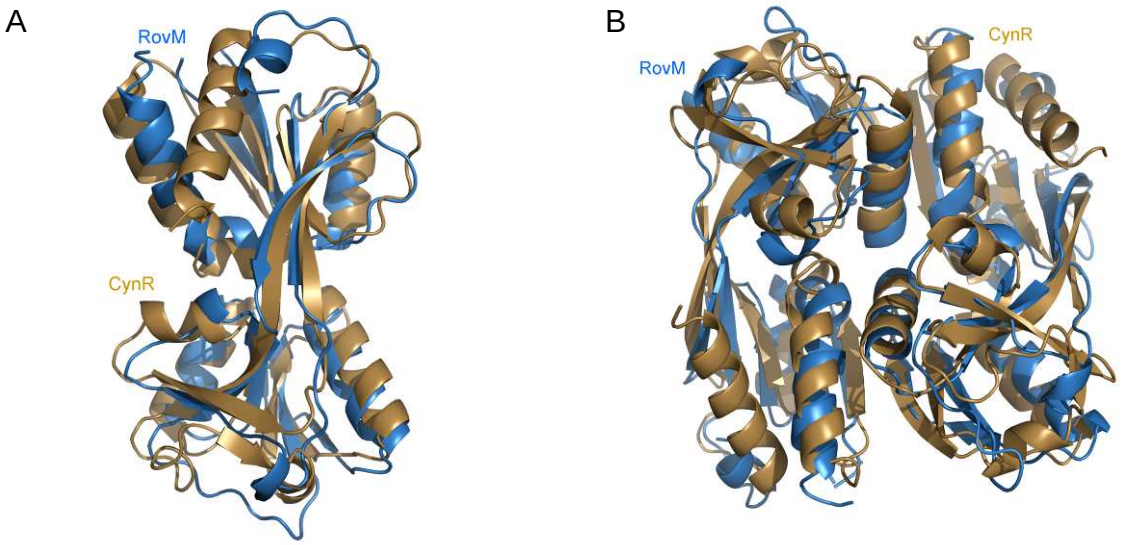


Fig. 67: Superposition of RovM-EBD with CynR. RovM-EBD is shown in blue, CynR in gold. A: monomers; B dimers.

An alignment of the sequences of RovM and its homologues CynR (*E. coli)* (Knapp & Hu 2009), CbnR (*Ralstonia eutropha)* (Muraoka *et al.* 2003) and BenM (*Acinetobacter baylyi)* (Ezezika *et al.* 2007) shows a high degree of conservation within the DBDs, indicating a conserved DNA binding mechanism (Fig. 68). In contrast, the EBD has only few conserved residues, mostly hydrophobic residues in the core of the protein. In particular, the C-terminal region is very dissimilar between the homologues. This distribution of conserved residues has also been observed for other classes of transcription factors such as the MarR family, which have a mostly conserved DBD with only few mutations to adapt to new promoter sequences and a weakly conserved ligand binding domain that allows binding of very different inducer molecules (Wu *et al.* 2003).

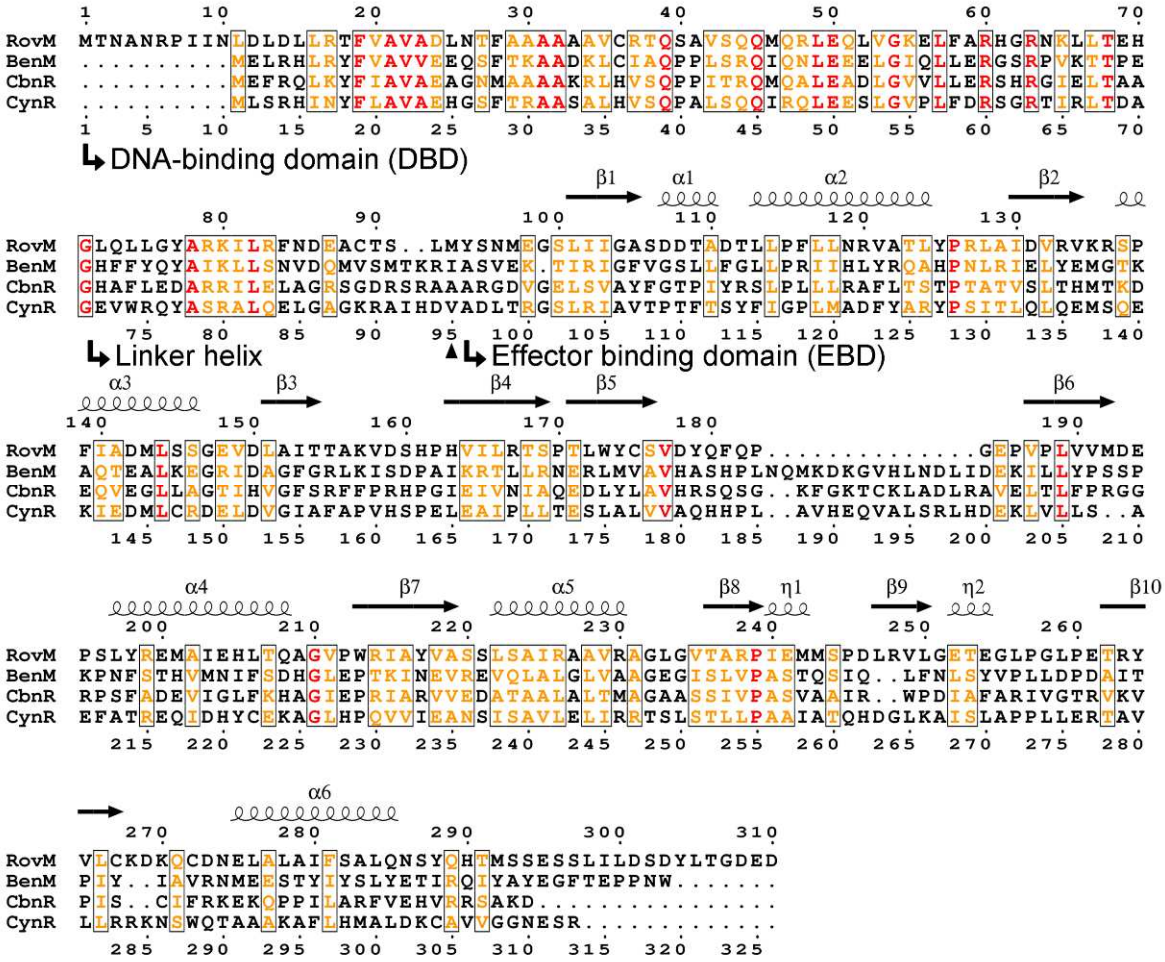


Fig. 68: Structure based sequence alignment of RovM with other LTTR proteins. The sequences of RovM (*Y. pseudotuberculosis*), BenM (*Acinetobacter baylyi*, pdb code 2f7a), CbnR (*Ralstonia eutroph*a, pdb code 1ixc) and CynR (*E. coli*, pdb code 3hfu) were aligned with ClustalW (Larkin *et al.* 2007) and displayed with ESPript2.2 (Gouet *et al.* 1999).

Of the known LTTR structures, full length RovM has the highest sequence identity to CbnR (20.4%; Fig. 68). A superposition of RovM with the structure of a monomer of the full-length CbnR shows that the EBD-Is align quite well, whereas the EBD-IIs seem to be tilted with respect to the EBD-I (Fig. 69A) (Muraoka *et al.* 2003). In the full-length RovM the DBD would be at the N-terminus of the protein, connected to the EBD *via* a long linker helix. The DBD and the linker helix could easily adopt a conformation very similar to that of CbnR as there are no clashes in this alignment.

Full length CbnR is a tetramer, which was also shown in this study to be true for full length RovM. The RovM-EBD on the other hand is a dimer in solution. Superposition of the RovM-EBD BB´-dimer with the CbnR tetramer shows that the RovM-EBD dimer is arranged in the same way as the dimers of the EBDs of CbnR (Fig. 69B). The linker helices and DBDs could be oriented in the same fashion as in CbnR

allowing DNA binding at two distant sites and DNA bending, which has been shown for RovM (Heroven & Dersch 2006).

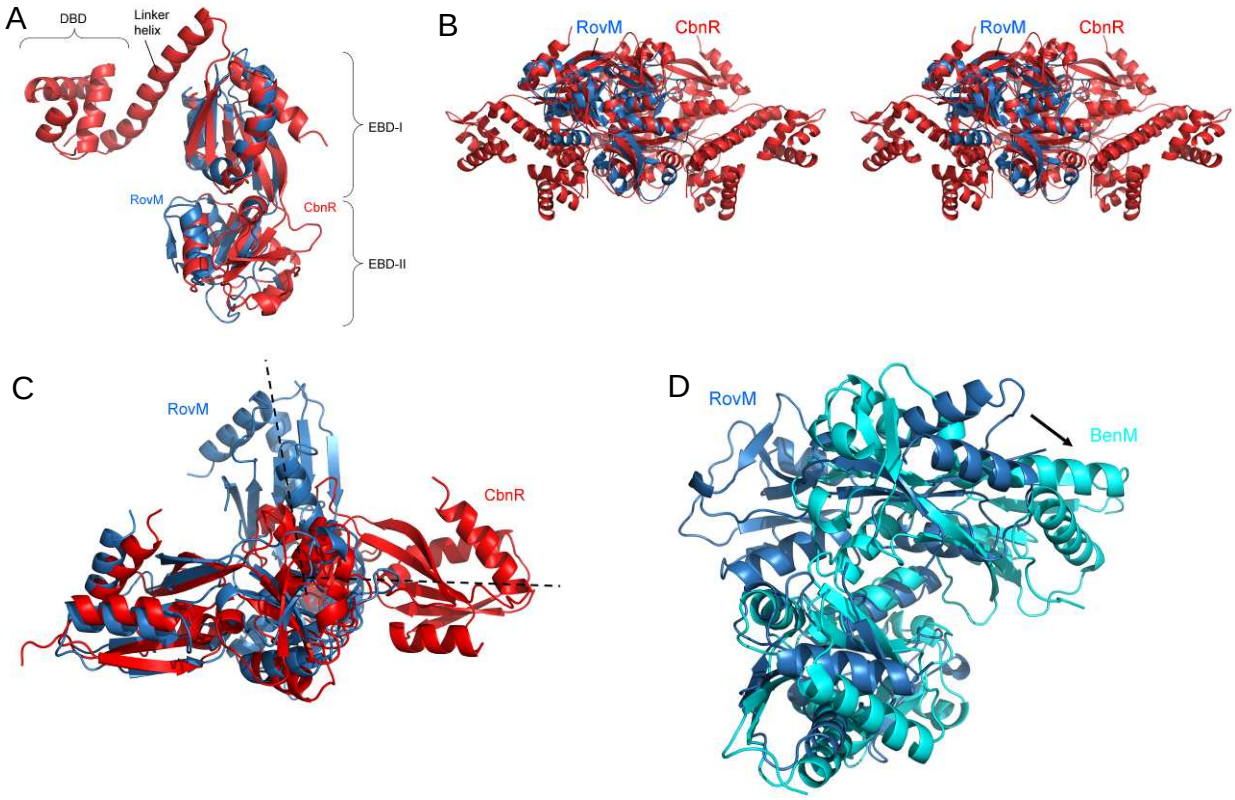


Fig. 69: Superposition of the structures of RovM-EBD and CbnR and BenM. RovM-EBD is shown in blue, CbnR in red. A: monomers; B: Stereo image of RovM-EBD dimer and CbnR tetramer; C: Superposition of two RovM and CbnR EBDs connected via the tetramer interface showing the resultant angle between the subunits. D: Superposition of RovM and BenM (cyan) connected via the tetramer interface showing a shift of the subunits between RovM and BenM.

In CbnR the interaction between the two dimers is mainly mediated by helix V of one subunit with the same helix of another subunit, related by a two-fold axis (Muraoka *et al.* 2003). This kind of interaction, called tetramer interface, has also been detected in other full-length LTTR structures, namely BenM (Ruangprasert *et al.* 2010) and DntR (Smirnova *et al.* 2004). A similar interaction is also present in the RovM-EBD structure called AA`` interface. Here the interactions are mostly mediated by helix 3 (which corresponds to helix V in CbnR) and helix 2. Yet, this interface varies considerably between the different structures. This is easily visualized by the angle between the two symmetry related helices 3 and 3`. In the CbnR structure the two helices are approximately orthogonal to each other (86°), whereas the angle is 129° for DntR and 152° for BenM (Ruangprasert *et al.* 2010). In contrast, in the RovM-

EBD structure these two helices are nearly parallel with an angle of 172°. Thus, the two EBDs related by the tetramer interface are roughly parallel to each other in CbnR, whereas they are nearly orthogonal in RovM-EBD (Fig. 69C). The structure in which the angle is closest to RovM-EBD is BenM, where the EBDs are also nearly parallel. But, in comparison with RovM-EBD, in BenM the EBDs are also shifted relative to each other (Fig. 69D). This shows that even though several LTTRs use the same interface, the resulting oligomeric structures are quite different.

To determine whether it would be possible for RovM-EBD to form a tetramer similar to that of CbnR (two pairs of EBDs connected via the tetramer interface between one EBD from each pair), four B subunits were picked accordingly (Fig. 70). Yet, in such an orientation the N-termini of two subunits were positioned very close to each other so that the linker helices in a full-length protein would probably clash. It thus seems unlikely that the full-length RovM tetramer would form in this manner. In the full-length RovM the strong interaction between the linker helices would probably dominate over the weak tetramer interaction and there might be no interaction between the EBD pairs as seen in 2ESN. On the other hand, the strong EBD interface in the asymmetric unit of RovM-EBD might also play a role in forming the tetramer or in cooperative binding of several RovM tetramers.

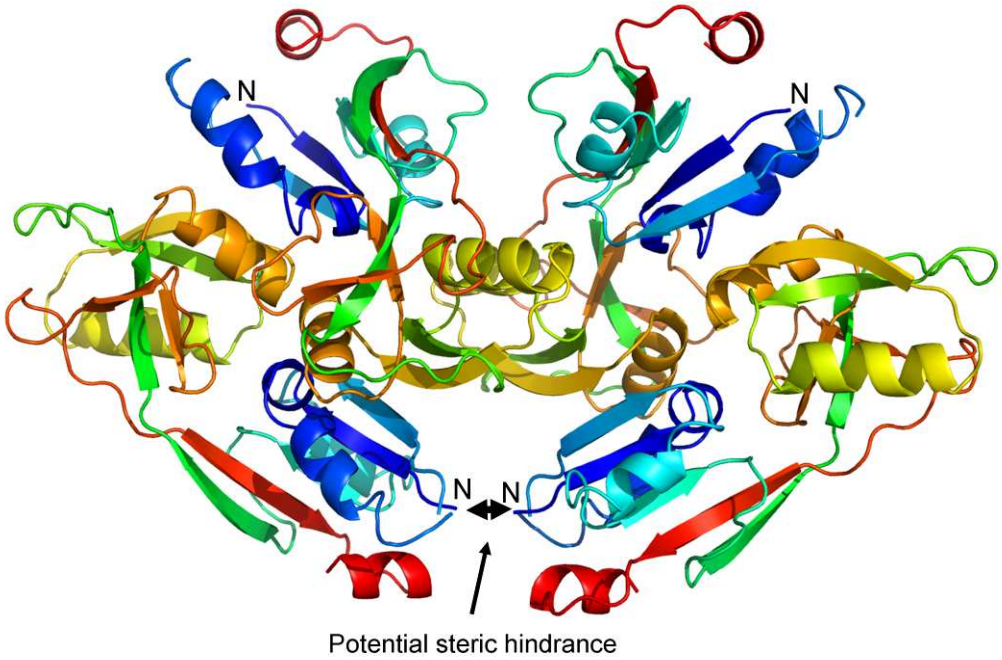


Fig. 70: Tetrameric assembly of RovM in the crystal packing similar to the tetramer of CbnR. The black arrow indicates a potential clash of the full-length RovM if oriented in this fashion.

5A.2.3 Analysis of inducer binding sites in RovM

For several LTTR proteins inducers have been described that influence the gene regulation of these proteins. For some of these proteins such as BenM it was possible to solve the structure in complex with its inducer (Ezezika *et al.* 2007). The inducer molecule (in this case *cis*, *cis*-muconate) was shown to bind in the cleft between EBD-I and EBD-II and binding resulted in a tilting of the two domains towards each other. This movement is thought to be relayed to the DBDs via the long linker helices and thus influence DNA binding. BenM was also shown to contain a secondary ligand binding site in EBD-I between helix 1 and sheet 4, where a benzoate molecule was bound (Ezezika *et al.* 2007). Fig. 71A shows the positions of both ligands after alignment with the RovM structure. The inspection of the primary binding site shows that there is a cavity in RovM, which is lined with few hydrophobic and several hydrophilic residues, some of which are similar in BenM (Fig. 71B). Yet, the cavity is much shallower than in BenM indicating that RovM would bind a smaller ligand than BenM.

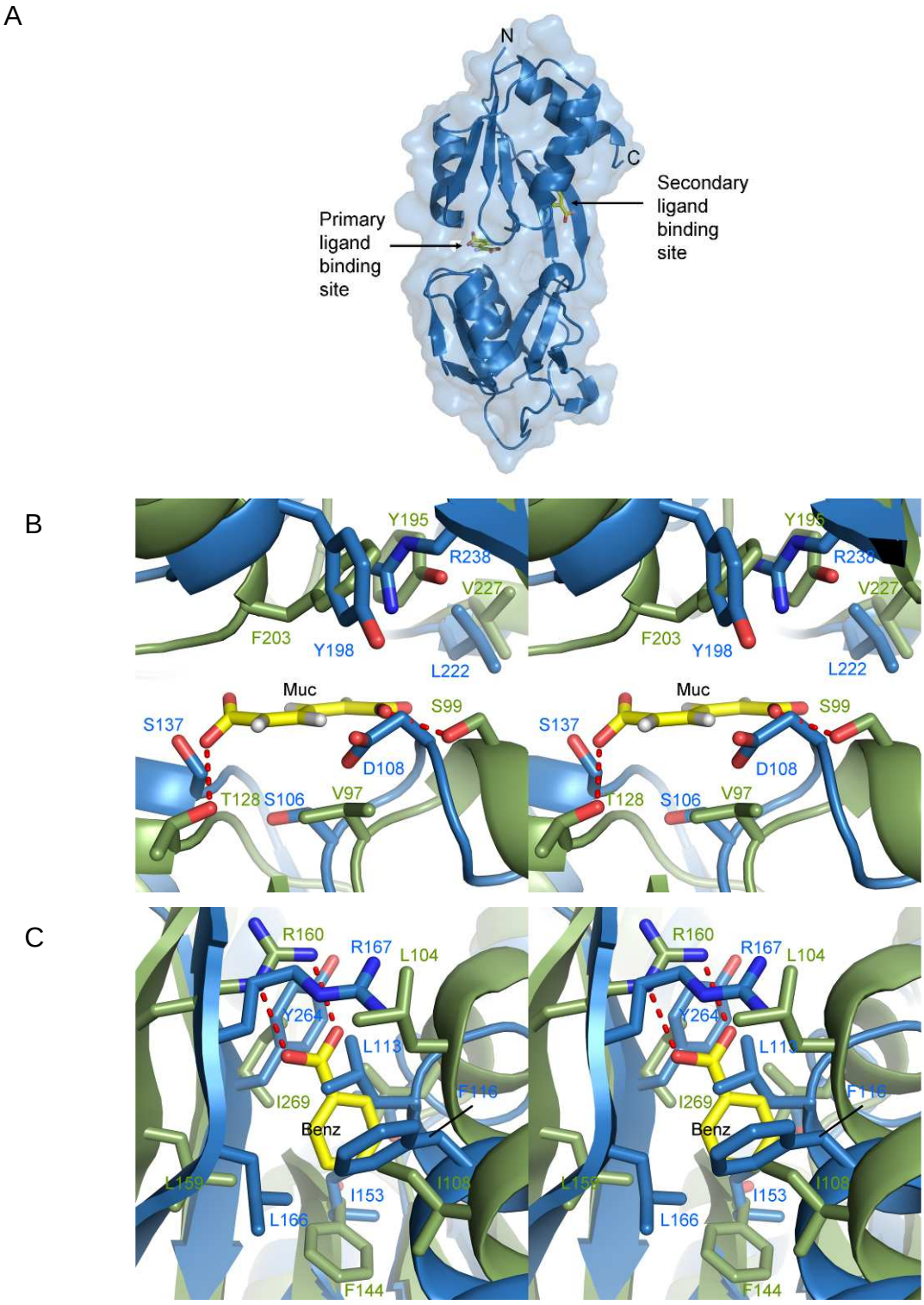


Fig. 71: Comparing BenM (green) ligand binding sites with the structure of RovM-EBD (blue). BenM ligands cis, cis-muconate and benzoate, are both shown as yellow sticks A: Cartoon and surface display of RovM-EBD. B: Stereo view: comparison of primary binding site residues between RovM-EBD and BenM. C: Stereo view: comparison of secondary binding site residues between RovM-EBD and BenM.

A cavity for the secondary binding site on the other hand is not visible for RovM. The ligand clashes with some side chains such as L113 and L166 and even the backbone of helix 1 (Fig. 71). Yet, the residues that form the secondary binding site in BenM are very similar to those in RovM. Thus, only after conformational changes in RovM might a secondary binding site similar to that observed in BenM become accessible. Future work will concentrate on finding inducer molecules for RovM and analysing their role in the regulation of invasin and other *Yersinia* virulence factors.

5A.2.4 Comparison between RovM from pathogenic *Yersinia* species

Y. pseudotuberculosis and *Y. enterocolitica* are related pathogens, which are both taken up via the faecal-oral route and cause similar diseases. Both use the same regulatory system including RovA and RovM to regulate the expression of virulence factors such as invasin. Even *Y. pestis*, a vector-borne pathogen which enters its host via a completely different route and causes different diseases, relies on this system for virulence factor regulation. The sequence of RovM is fully conserved between *Y. pseudotuberculosis* and *Y. pestis*, whereas there are some differences between *Y. pseudotuberculosis* and *Y. enterocolitica* (Fig. 72).

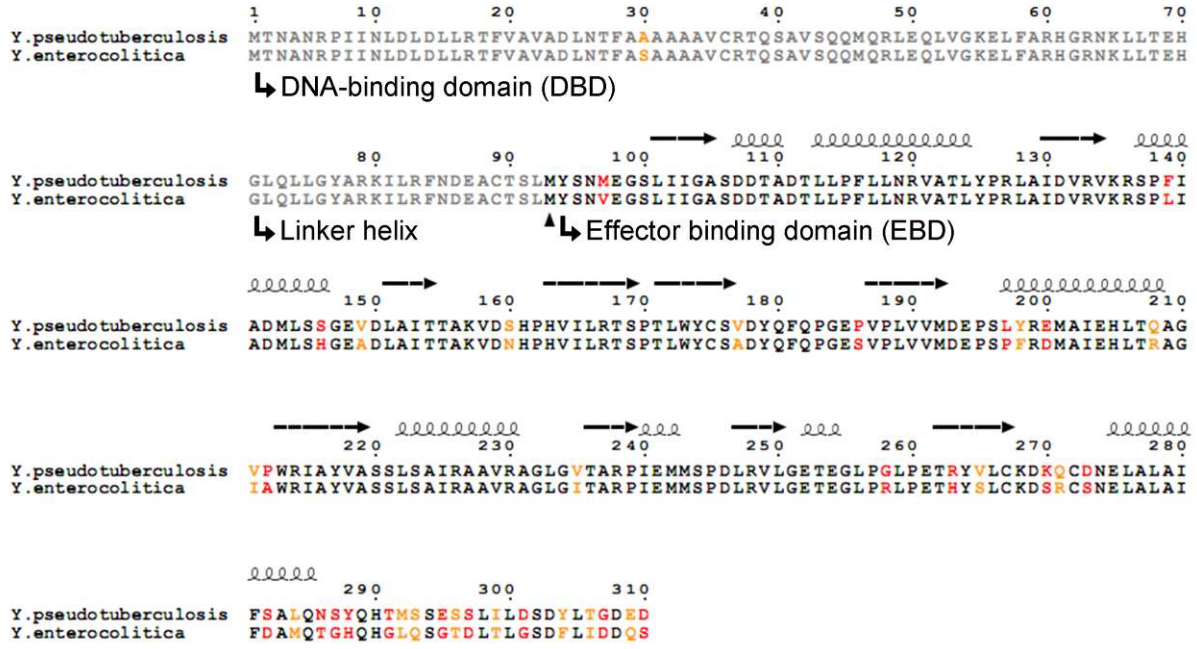


Fig. 72: Structure based sequence alignment of RovM from *Y. pseudotuberculosis* and *Y. enterocolitica*. The sequences were aligned with ClustalW (Larkin *et al.* 2007) and displayed with ESPript2.2 (Gouet *et al.* 1999)**.**

Few mutations are found in the DBD, which shows a conserved DNA binding mechanism, and the first 45 amino acids of the EBD. Most of the mutations can be found at the end of helix 7 and the following region, which is probably unstructured and was not detectable in the electron density due to its flexibility. Mapping the changed residues onto the structure of RovM demonstrates that most of the mutations can be found on the surface of the protein (Fig. 73). Yet, two mutations (L197P and Y198F) are located within the potential inducer binding site between EBD-I and EBD-II. So far, it is unclear what the inducer of RovM is or whether there is one at all. But since RovM has been shown to mediate invasin expression in response to nutrient availability, it is conceivable that some small molecule connected to metabolism may act as an inducer for RovM. Whether the mentioned mutations in the ligand binding pocket affect the binding selectivity or affinity of the putative ligand will be subject of further investigation. RovA and RovM genes have also been found in virtually all other *Yersinia* species with the most distant RovM from *Y. intermedia* sharing 88% sequence identity. Yet, these other species are non-pathogenic in humans and have no invasin gene. Probably, the whole RovA/RovM regulatory system originally performed other functions and has evolved in pathogenic species to regulate virulence genes as well.

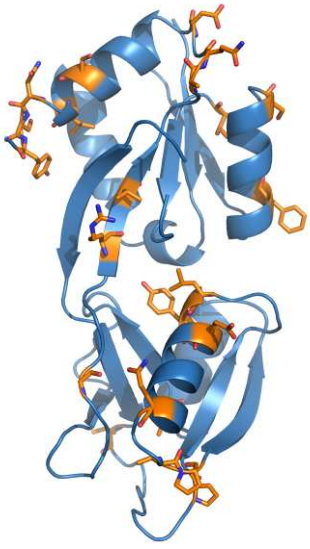


Fig. 73: Differences in RovM between *Y. pseudotuberculosis* and *Y. enterocolitica*, shown as orange sticks.

In conclusion we have presented the structure of RovM-EBD from *Y. pseudotuberculosis* as well as analysis of the oligomeric state of RovM-EBD and full

length RovM. RovM-EBD exists as a dimer, while full length RovM probably assembles into a tetrameric dimer of dimers like the homologue CbnR. Additionally, a cavity was detected in RovM-EBD which could serve as a binding site for a small inducer molecule.

5B Polyketide synthases

5B.1 The octenoyl-CoA carboxylase/reductase CinF

5B.1.1 CinF generates an unusual building block for polyketide synthesis

Many drugs that are in use today, for example as antibiotics or anti-cancer drugs, are derived from natural products that are produced by polyketide synthases (PKS). These huge multi-enzyme biosynthesis clusters use small building blocks, called extender units, to build large, complex molecules with biological activity. A typical extender unit is methylmalonyl-CoA that is produced from crotonyl-CoA by so called crotonyl-CoA carboxylase/reductases (CCR) via reductive carboxylation. However, other extender units have been described such as hexylmalonyl-CoA and chloromethylmalonyl-CoA. In order to understand the molecular basis of the reaction mechanism as well as the substrate specificity, in this work we have characterised biochemically as well as structurally CinF that generates the unusual extender unit octenoyl-CoA.

5B.1.2 CinF is a tetramer

Here we describe the first structure of a carboxylase/reductase. Size exclusion and PISA analysis suggest that CinF forms a tetramer. The closest homologous structures of CinF, two putative crotonyl-carboxylases/reductases (CCR, pdb codes: 3HZZ and 3KRT, both with 51% sequence identity) which are both yet unpublished, also appear to form a similar tetramer. Tetrameric assemblies have also been shown for some other related protein such as SsADH (1R37, 24% sequence identity) (Esposito *et al.* 2003), whereas others such as PIG3 (2J8Z, 25% sequence identity) (Porte *et al.* 2009) are dimeric. These dimers correspond to the strong dimer interface observed in CinF, which creates the continuous β-sheet between the cofactor binding domains.

5B.1.3 CCRs contain several insertions compared to other reductases

Comparison of CinF to homologous proteins such as PIG3 reveals some significant differences between the otherwise similar structures (Fig. 74A). An insertion of several amino acids in the cofactor binding domain of CinF leads to an elongation and twist by 31° of helix α14 compared to PIG3. This helix is involved in binding the CoA of the dimer partner and the elongation and twist places the respective residues (R286 and R293) in proximity to the CoA. Another insertion in CinF exists between helix α5 and sheet β5, which forms two additional helices α6 and α7 as well as another short sheet β4, whereas in PIG3 there is only a short loop. This insertion also plays an important role in substrate binding as the CoA is wound around helix α6 and held in place by several interactions with it. Another short insertion involved in CoA binding can be found after sheet β1, leading to the formation of a helix (α3). Additionally, CinF contains additional amino acids at its N- and C-terminus, which are forming two helices each. Yet, the role of these additional regions is not clear, as they are neither involved in ligand binding nor tetramer formation.

The three structures of carboxylases/reductases are overall very similar, with an rmsd of 1.3 Å and 1.5 Å for common C_α atoms of chain A when CinF was compared with the CCRs from *S. coelicolor* (3KRT) and *S. collinus* (3HZZ), respectively (Fig. 74B). These two proteins are proposed to catalyse the reductive carboxylation of crotonyl-CoA, whereas CinF uses octenoyl-CoA, which has a longer acyl-chain. We have investigated the molecular basis for the difference in substrate specificity. Superposition of the structures of CinF and the CCRs shows an elongated hydrophobic pocket for the binding of the octenoyl-chain in CinF, whereas in the other carboxylases/reductases no such pocket was detectable. Instead the potential pocket is blocked by the large residues Phe370 and Ile171 in 3KRT while the corresponding small residues in CinF are Gly362 and Ala163 (Fig. 74C). Thus, these two positions seem to determine the substrate specificity in carboxylases/reductases.
A putative CCR from *S. hygroscopicus* (Accession Nr. AAR32675), which has 90% sequence identity with CinF, is probably also able to use octenoyl-CoA as a substrate as none of the changed amino acids are involved in ligand binding (Fig. 75). It also contains a glycine and an isoleucine at the same positions as CinF where these determine the substrate specificity.

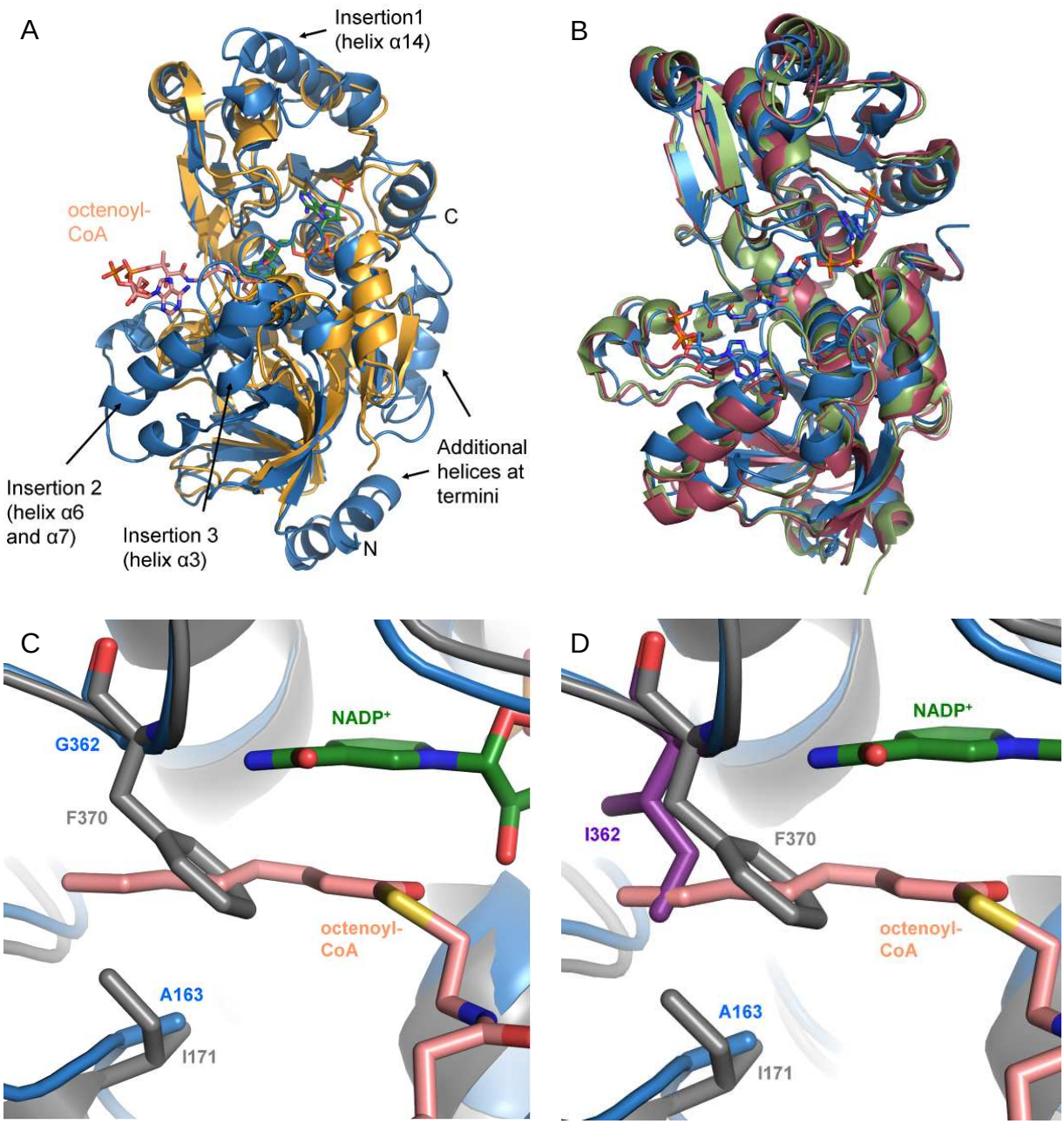


Fig. 74: Comparison of CinF with other structures. A: Comparison of CinF (blue) with PIG3 (yellow) showing the insertions of CinF. B: Superposition of CinF (blue) with CCR from *S. collinus* (3KRT, green) and *S. coelicolor* (red), the ligands of CinF are shown as blue sticks. C: Comparison of the substrate binding sites of CinF (blue) and the CCR from *S. collinus* (3KRT, grey). The binding site of CinF is lined by small residues (G362 and A163), whereas this binding site is blocked by I171 and F370 in the CCR from *S. collinus*. This explains why CinF is able to use octenoyl-CoA as a substrate whereas the other CCRs are not. D: In SalG instead of the glycin residue there is an isoleucin residue. This residue was modelled into the ligand binding site of CinF (purple). It explains why SalG is able to use the longer Cl-crotonyl-CoA instead of the usual crotonyl-CoA.

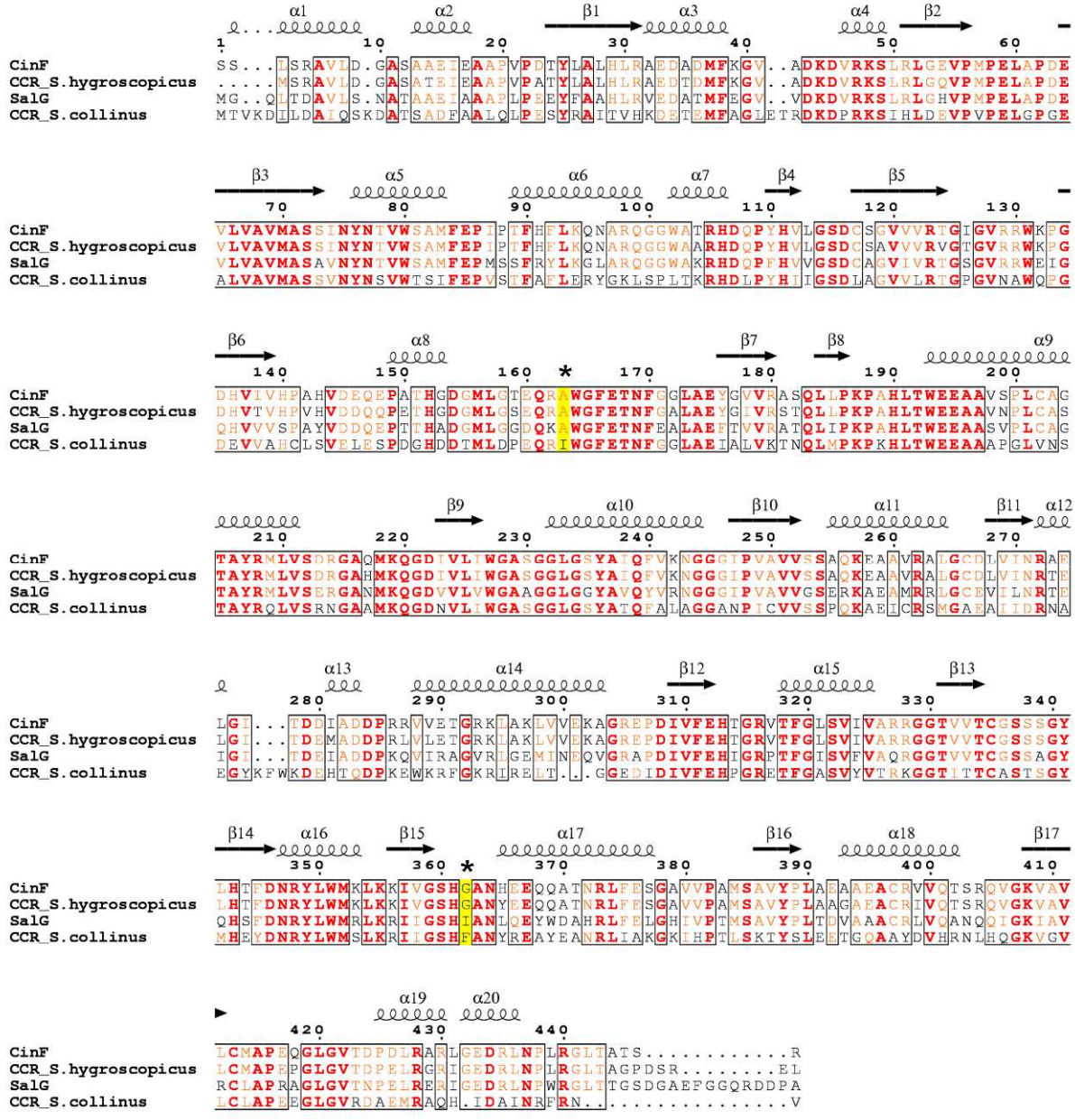


Fig. 75: Structure based sequence alignment of CinF with other carboxylase/reductase proteins. The sequences of CinF (*Streptomyces coelicolor*), a putative octenoyl-CoA carboxylase/reductase (*Streptomyces hygroscopicus*), SalG (*Salinospora tropica*) and a putative CCR (*Streptomyces collinus*, pdb code 3KRT) were aligned with ClustalW (Larkin *et al.* 2007) and displayed with ESPript2.2 (Gouet *et al.* 1999). The positions of the residues determining the substrate specificity are marked in yellow and with an asterisk.

Another protein that has been shown to recognize a different substrate than crotonyl-CoA is SalG from *Salinispora tropica* which is involved in the production of salinosporamides (Eustaquio *et al.* 2009). SalG has a 7-fold higher catalytic

efficiency with chlorocrotonyl-CoA than with crotonyl-CoA. SalG shares 70% sequence identity with CinF and a comparison of the two positions shown before to be important for substrate specificity indicates a different substrate binding pocket for this protein as well (Fig. 75). SalG also has an alanine in the same position as CinF, Ala163, but instead of the Gly362 in CinF or the Phe 370 in 3KRT in has an isoleucine. *In silico* modelling of this mutation into the otherwise unchanged ligand binding pocket of CinF allows placement of the isoleucine residue without major clashes in one rotamer (Fig. 74D). In this orientation the isoleucine reduces the size of the substrate binding pocket of CinF so that instead of octenoyl-CoA only pentenoyl-CoA will fit into the pocket. This agrees well with the data of Eustaquio *et al*. (2009) that chlorocrotonyl-CoA is the native substrate which has the same size as pentenoyl-CoA. To strengthen this hypothesis two mutants of CinF, G362F and G362I, have also been characterised by Liujie Huo (HIPS, Saarbrücken). These mutants were unable to use octenoyl-coA which underlines the importance of this residue. It thus seems, that by varying the size of the substrate binding pocket different substrates can be accommodated which leads to different substrate specificities of these enzymes. These findings are highly interesting for the generation of engineered CCRs to produce drugs incorporating uncommon extender units.

5B.1.4 Reaction mechanism

It has been demonstrated that CCRs carboxylate the substrate on the *re* side to yield a product with (2S)-stereochemistry, while the hydride transfer from the NADPH is pro-(4R) specific to the *re* face of the substrate (Erb *et al.* 2009). The arrangement of the substrates in the CinF structure clearly agrees with these findings suggesting a common evolutionary origin of carboxylase/reductases with medium-chain reductases. However, CCRs have evolved to bind CO_2 and carboxylation is preferred over sole reduction of the substrate (Erb *et al.* 2009). Unfortunately, it was not possible to detect any electron density for CO_2 in the CinF crystals, even after aeration with CO_2 at high pressure. Therefore, *in silico* docking was used to identify a potential CO_2 binding site. In this binding site, the CO_2 is bound by hydrogen bonds to a glutamate and an asparagine residue and by hydrophobic interactions with a phenylalanine (Fig. 76A). The hydrogen bonds would increase the electrophilicity of the molecule, thus making it more susceptible for the nucleophilic attack from the

double bond of the octenoyl-CoA. The hydrophobic phenylalanine, on the other hand, probably restricts the binding of water in this pocket to prevent the reduction of the octenoyl-CoA without carboxylation.

A few other proteins have been crystallized in complex with CO_2, such as human carbonic anhydrase (pdb code: 2VVA) (Sjoblom *et al.* 2009), 2-ketopropyl coenzyme M oxidoreductase/carboxylase (2-KPCC, *Xanthobacter autotrophicus*, pdb code: 3Q6J) (Pandey *et al.* 2011), phosphoenolpyruvate carboxykinase (PCK, 2PXZ, *E. coli*) (Cotelesage *et al.* 2007) and an Rh protein (pdb code: 3B9Z, *Nitrosomonas europaea*) (Li *et al.* 2007). In most structures the CO_2 is located in a hydrophobic pocket, but bound by various binding partners such as backbone nitrogens (carbonic anhydrase) or arginines, lysines and tyrosines (PCK). Interestingly, in the Rh protein the CO_2 is located in a pocket that resembles the proposed binding site in CinF (Fig. 76B). The CO_2 is bound by an asparagine residue on one side and an aspartate on the other, while a phenylalanine residue is found below the CO_2. Additionally, there is a serine located in the vicinity. This supports that CO_2 is most likely bound in the proposed binding pocket in CinF.

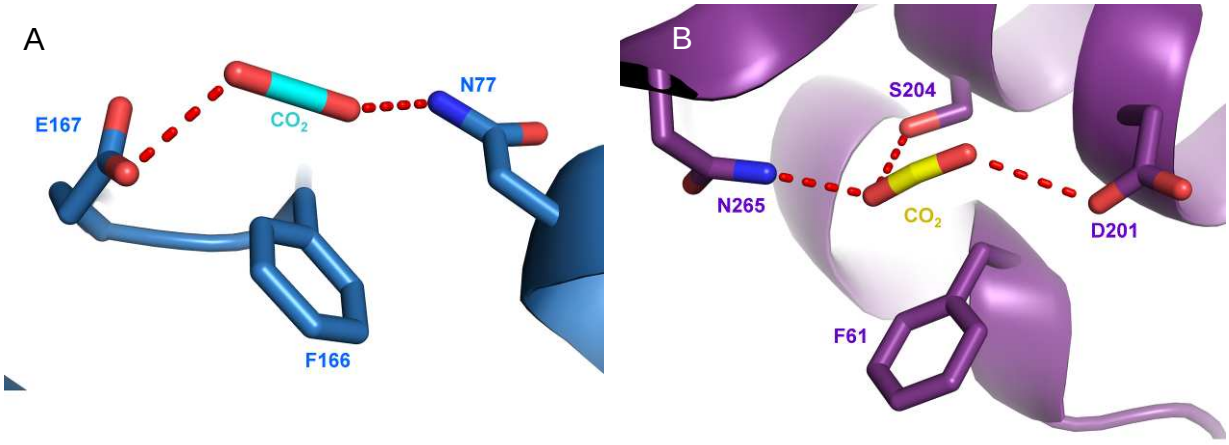


Fig. 76: CO_2 binding by different proteins. A: Proposed CO_2 binding pocket of CinF. B: CO_2 binding pocket of Rh protein.

5B.2 The thioesterase SpirTE

5B.2.1 SpirTE causes the release of a linear product

PKS biosynthesis relies on the covalent attachement of the growing product via a thioester bond to the Ppant-arm of the ACP domains of the PKS. After the last step of synthesis the final product has to be released from the PKS, usually by a TE domain that cleaves the thioester bond and often also catalyses the cyclisation of the product. SpirTE is the terminal domain of the spirangien-PKS cluster and causes the release of a linear product instead of a typical cyclised product. To understand the mechanism of the reaction the structure of SpirTE was solved, revealing a narrow substrate tunnel in which the active site is located.

5B.2.2 The narrow substrate tunnel of SpirTE inhibits product cyclisation

SpirTE shares the same overall fold with all PKS TE domains characterised so far (Fig. 78). A superposition with DEBS TE (36% sequence identity) (Tsai *et al.* 2001), PIK TE (28% sequence identity) (Akey *et al.* 2006) and TMC TE (32% sequence identity) (Scaglione *et al.* 2010) results in an r.m.s.d. of 1.6 Å, 1.7 Å and 1.9 Å, respectively, for common C_α atoms. SpirTE and TMC TE both catalyse the release of a linear product while DEBS TE and PIK TE additionally catalyse a cyclisation of the product. Yet, the spirangien as well as the TMC precursor molecules, which are the substrates for their respective TE domains, contain additional hydroxyl groups which would, in principle, allow cyclisation. Thus, the capability to form cyclic products or only linear products has to be a consequence of the structure of the TE domain. Even though the overall structures are similar, the size of the tunnel within the proteins is quite different. The tunnels of DEBS TE and PIK TE, which both catalyse the cyclisation of the product, are substantially wider than the tunnels of SpirTE and TMC TE, which release a linear product. The tunnel is very constricted in TMC TE and SpirTE, especially at the centre of the tunnel, where the catalytic site is located. This probably prevents the TE-bound substrate from looping back onto itself for cyclisation. The structure of PIK TE has been solved in complex with a cyclic product analogon 10-deoxymethynolide and as a phosphopentaketide adduct (Akey *et al.* 2006). Superposition of these structures with SpirTE clearly shows that such a cyclic arrangement of the product would not be possible in SpirTE. The tunnel of SpirTE is

too narrow for such a product mainly due to the larger residues S151 and Y87 in SpirTE (G149 and A78 in PIK TE). In TMC TE the tunnel is constricted by S133, V67 and additionally F182.

It is therefore feasible that cyclisation of the spirangien and TMC precursor is prohibited by this bottleneck created by bulky amino acids around the active site.

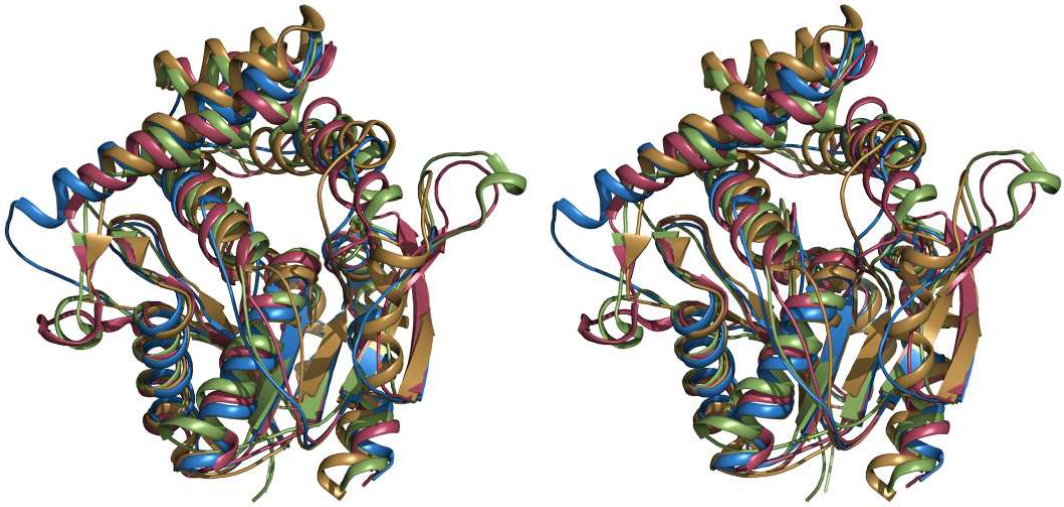


Fig. 77: Comparison of SpirTE and other TE structures. Stereo image of superposition of PKS TE domains. Blue: SpirTE, red: DEBS TE, green: PIK TE, ochre: TMC TE

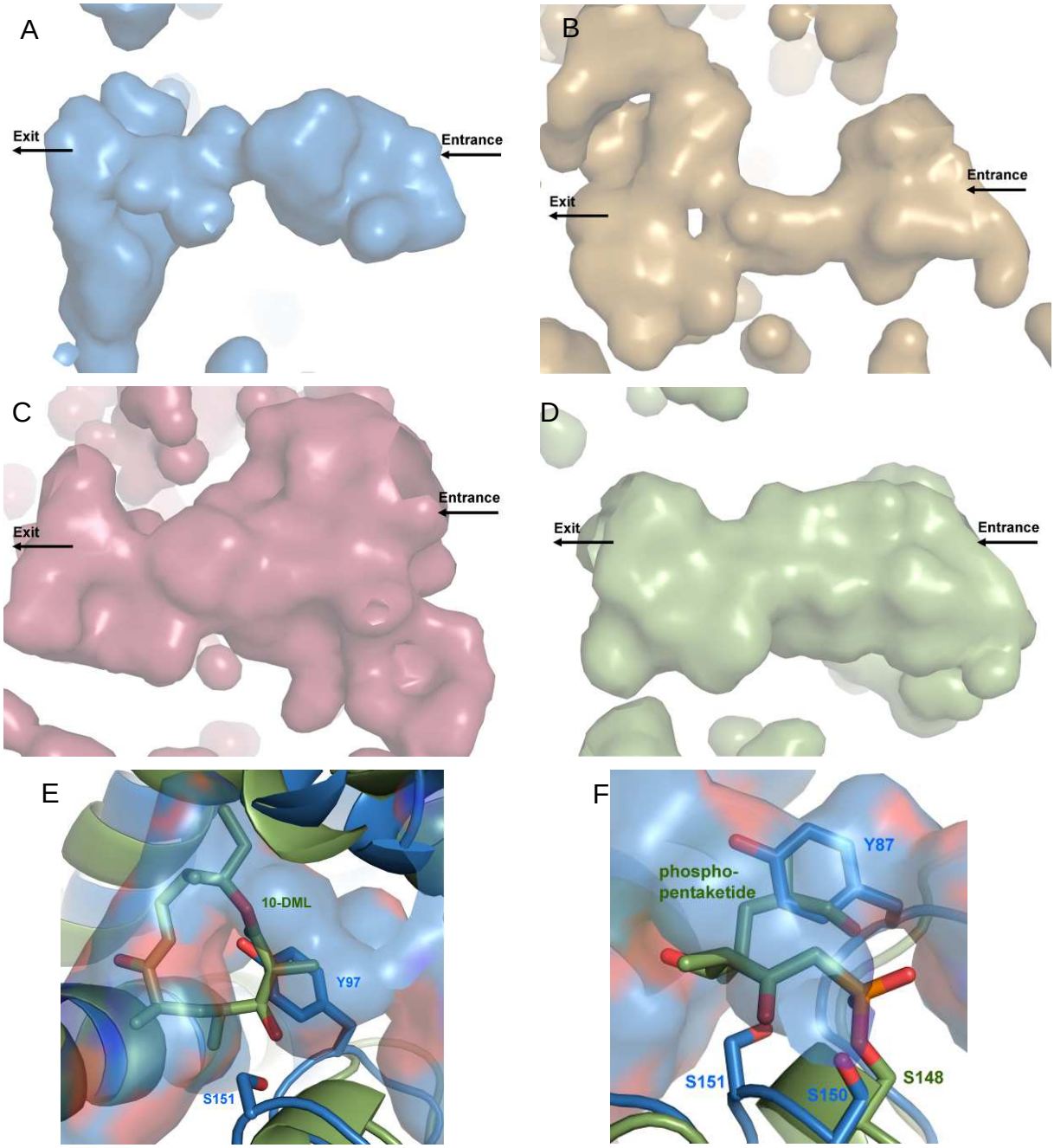


Fig. 78: Comparison of SpirTE and other TE structures. A-D: Inner surface of substrate tunnel of TE domains (B: SpirTE, C: TMC TE, D: DEBS TE, E: PIK TE). F-G:Alignment of SpirTE (shown as cartoon and inner surface) with PIK TE in complex with its model product 10-deoxymethynolide (10-DML) (F) and as phosphopentaketide adduct (G).

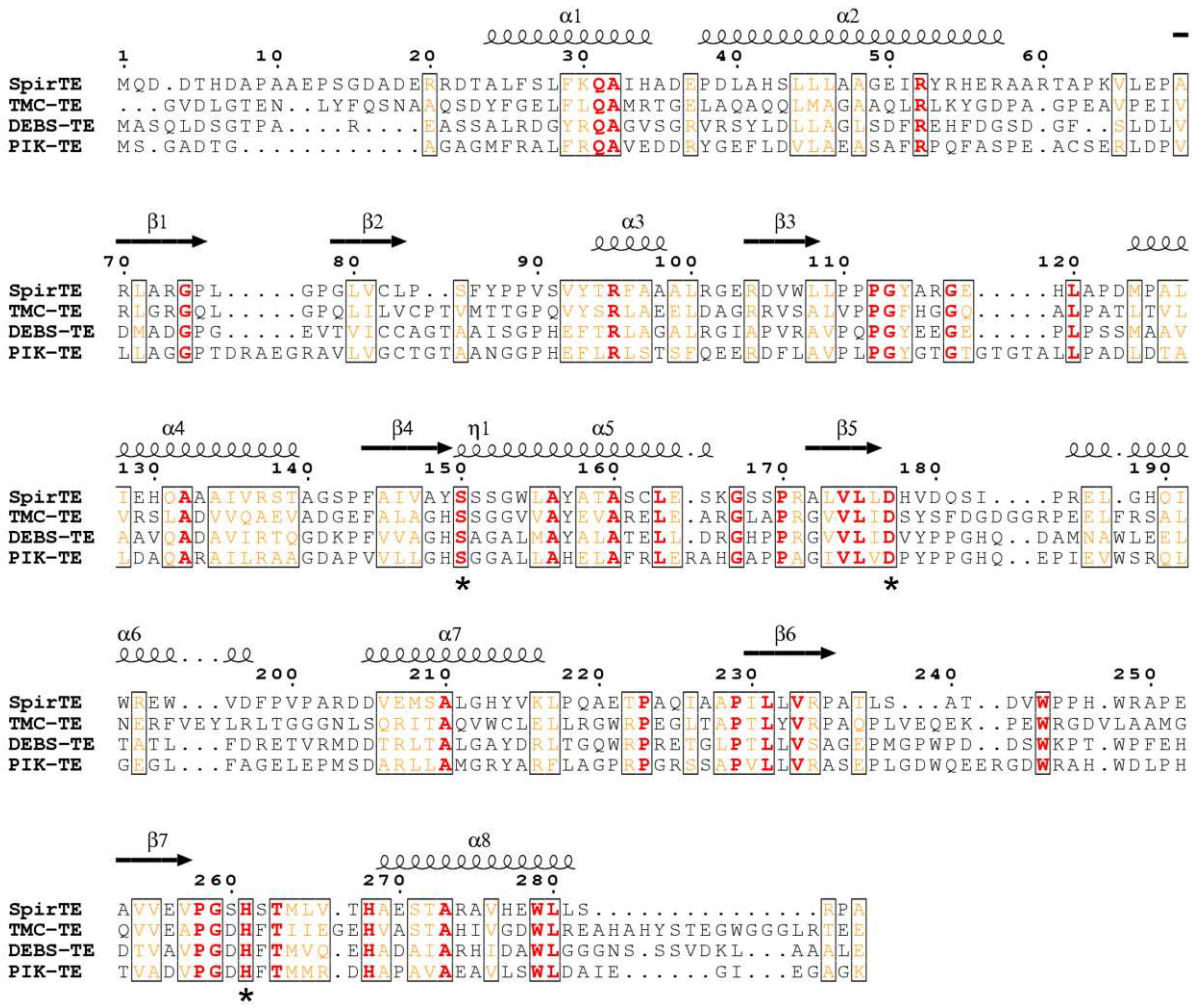


Fig. 79: Structure based sequence alignment of SpirTE with other PKS TE domains. The sequences of SpirTE (*S. cellulosum*), TMC-TE (*Streptomyces sp. ck4412*, pdb code 3lcr), DEBS-TE (*Saccharopolyspora erythraea*, pdb code 1kez) and PIK-TE (*Streptomyces venezuelae*, pdb code 2h7x) were aligned with ClustalW (Larkin *et al.* 2007) and displayed with ESPript2.2 (Gouet *et al.* 1999). The residues of the catalytic triade are marked by an asterisk.

Another interesting feature is visible when comparing the sequences of the TE domains. Normally, the catalytic serine residue is found within the signature motif GxSxG such as GHSSG in Tautomycetin TE (TMC TE) (Scaglione *et al.* 2010). This leads to the formation of a so called catalytic elbow which makes the catalytic serine stand proud on the active site surface (Ollis *et al.* 1992). Surprisingly, the corresponding sequence in SpirTE is AYSSS, where the two glycines are replaced by the larger amino acids alanine and serine (Fig. 79). This leads to an unwinding of the first turn of the following α-helix α5 (Fig. 80). Whether this has any impact on the catalytic activity is not clear. A comparative activity assay has been performed with SpirTE and DEBS TE with p-nitrophenyl esters as substrates (Buntin *et al.* 2010b) and showed that the activity of SpirTE markedly increases with the substrate chain

length, while DEBS TE seems to be inhibited by the short chain substrates. Only with the pentanoyl ester could an apparent specificity constant be determined for DEBS TE which was about one third of the constant for SpirTE. These results indicate that SpirTE prefers longer aliphatic substrates as would be expected, yet it is improbable that the slight change in active site geometry can account for the observed difference in activity between SpirTE and DEBS TE.

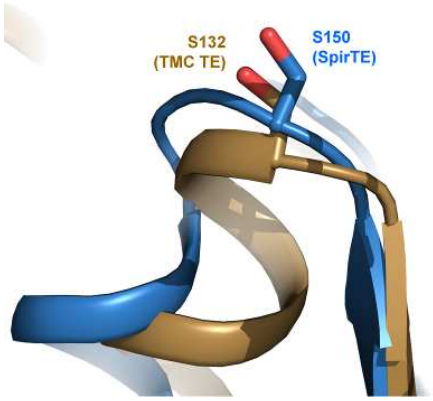


Fig. 80: Comparison of the catalytic serine in SpirTE (blue) and TMC TE (ochre). Due to changes in the catalytic site signature motif in SpirTE the nucleophilic elbow is not as pronounced as in other TE domains.

6 Conclusions and Outlook

6A Virulence regulation in pathogenic *Yersiniae*

RovA is the central player in a complex regulatory network involved in virulence factor regulation in pathogenic *Yersiniae* (Heroven & Dersch 2010). In this work the role of RovA has been elucidated biochemically and structurally. Several small molecules such as salicylate were found to bind to RovA with millimolar affinity. Further experiments revealed that binding of salicylate reduces the temperature dependent conformational changes observed by CD spectroscopy as well as the DNA binding affinity. Yet, to date it is not clear whether there really is an inducer for RovA. Future experiments such as mass spectrometrical analysis of capturing experiments from cell extracts might shed some light into the regulation of RovA by small molecules.

Furthermore, a mutant of RovA (P98S) was discovered in certain *Y. enterocolitica* strains that was more stable *in vivo* than the wild type. Experiments were undertaken to characterize this mutant in vitro. However, it turned out that the mutant behaved exactly like the wild type and it was thus concluded, that the mutation probably had an impact on the recognition of RovA by proteases. Two other mutants (G6F and A10F) were designed to block the ligand binding pocket to gain insight into the role of ligands on the regulation by RovA. While G6F behaved like the wild type, A10F was significantly more stable *in vitro* and had lower ability to bind salicylate. Further studies will be carried out to investigate the impact of these mutations *in vivo*.

It was possible to solve the structure of RovA-apo, in complex with DNA as well as in complex with the putative inducer molecule salicylate. These structures show that RovA is very flexible and can probably adopt different conformations in solution alone and in complex with ligands. For example, binding of salicylate induces a closed conformation compared to the DNA bound form which would prevent DNA binding. In the structure in complex with salicylate, two salicylate molecules were found in every monomer of RovA. One is tightly bound in a hydrophobic pocket, whereas the other is bound on the surface to the DBD. Comparison with structures of other MarR-type

proteins showed that the binding site in the hydrophobic pocket is conserved and probably plays an important role in the regulation of these proteins. The other binding site, however, was only also observed in the close homologue SlyA from *S. typhimurium.* Therefore, the physiological relevance of this binding site is questionable. Further studies, for example by NMR, could help to clarify this issue and provide insights into the conformational flexibility of RovA in solution.

Additionally, the DNA bound structure revealed how RovA interacts with the DNA. Only few amino acids form specific interactions with DNA bases which probably explains the relaxed DNA sequence specificity observed for RovA compared to specific transcription factors like OhrR. It also shows that RovA binds to linear DNA and thereby possibly counteracts the DNA bending induced by its inhibitors H-NS and RovM. Another interesting question arose by comparison of RovA and its homologue SlyA from *S. typhimurium*. These two proteins share a sequence identity of 75% and in particular in the DBD these two proteins are virtually identical. Therefore, it is reasonable to assume they have the same DNA binding specificity. However, the DNA recognition sequence that has been determined for SlyA is distinct from that predicted for RovA. Further experiments have to be conducted to explain this puzzling discrepancy.

The structure of the effector binding domain of RovM (RovM-EBD), an inhibitor of RovA, has also been solved in this study. The structure exhibited a potential small ligand binding pocket, which has also been described for other LTTR proteins. However, so far no inducer molecule is known for RovM (Heroven & Dersch 2006). Therefore, a transposon screen for *Yersinia* strains that show differences in RovA expression might lead to the identification of enzymes that are involved in the production of an inducer for RovM. This would help to understand how RovA is regulated by RovM. Once an inducer has been identified for RovM one could try to cocrystallize it with RovM. The structure of RovM in complex with its inducer would give insights into the conformational changes that take place upon inducer binding. Additionally, it would be very interesting to solve the structure of full-length RovM in complex with DNA as very little is known so far about the DNA binding of LTTRs and no structural information is available on this subject.

6B Polyketide synthases

In this work the first crystal structure of a carboxylase/reductase (CinF) has been solved in complex with its ligands $NADP^+$ and octenoyl-CoA. The structure grants insights into the novel enzymatic reaction, the reductive carboxylation without biotin or metal ions as cofactors, as well as provides a molecular basis for the understanding of the substrate specificity of this class of enzymes. Future work should be directed towards the introduction of mutations into typical crotonyl carboxylase/reductases to test whether it is possible to enhance their substrate specificity towards longer substrates as in CinF. It might also be possible to further mutate CinF to generate more space for even longer or even branched substrates. Additionally, the CO_2 binding and activation by CinF should be target of future efforts as it could only be docked into the structure in this work. To this end, mutations in the amino acids that have been proposed to be important should be generated. Another worthwhile aim would be the closer investigation of TgaD, a protein that seems to catalyze the same reaction as CinF, yet, has only little sequence identity with it (Buntin *et al.* 2010a).

By changing the substrate specificity of carboxylase/reductases it might become possible to allow the incorporation of unusual extender units into polyketides. Especially the introduction of chlorinated extender units such as in cinnabaramide (Buntin *et al.* 2010a) and salinosporamide (Eustaquio *et al.* 2009) can make new modes of actions possible for polyketide derived drugs. This would hugely enhance the combinatorial potential to produce new drugs with improved properties such as reduced resistance or higher efficacy.

In order to be able to produce tailor-made drugs by changing PKS systems the TE domains are of importance. Several TE domains have been described that catalyze the cyclization of their products (Akey *et al.* 2006; Tsai *et al.* 2001). However, little is known about TE domains that release a linear substrate (Scaglione *et al.* 2010). In this work SpirTE, the terminal domain from the spirangien PKS biosynthesis cluster, has been characterized structurally. The structure showed a substrate tunnel in which the active site is located. Compared to substrate tunnels from cyclizing TE domains, the tunnel of SpirTE is much more constricted. This prevents the TE bound product from bending back onto itself in contrast to the findings in PIK TE where a

macrocycle is formed (Akey *et al.* 2006). A similar substrate tunnel geometry has also been observed for TMC TE, which also produces a linear product (Scaglione *et al.* 2010).

Future work will focus on the crystallization of an inactive mutant of SpirTE in complex with its substrate. The structure in complex with its substrate will give insights into how the substrate is recognized by the protein. Perhaps, by identifying and mutating residues which constrict the tunnel it might become possible to allow TE domains that only cleave off the product also to catalyze cyclization. This might allow the production of completely novel compounds.

Novel and optimized drugs are in strong need in these times of rising antibiotic resistance in bacteria and the accompanying re-emergence of pathogens that were believed to be defeated.

7 References

Achtman, M., K. Zurth, G. Morelli, G. Torrea, A. Guiyoule & E. Carniel. 1999. *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proc Natl Acad Sci U S A* **96**, 14043-8.

Akey, D.L., J.D. Kittendorf, J.W. Giraldes, R.A. Fecik, D.H. Sherman & J.L. Smith. 2006. Structural basis for macrolactonization by the pikromycin thioesterase. *Nat Chem Biol* **2**, 537-42.

Albers-Schoenberg, G., B.H. Arison, J.C. Chabala, A.W. Douglas, P. Eskola, M.H. Fisher, A. Lusi, H. Mrozik, J.L. Smith & R.L. Tolman. 1981. Avermectins. Structure determination. *Journal of the American Chemical Society* **103**, 4216-4221.

Alekshun, M.N. & S.B. Levy. 1999. The mar regulon: multiple resistance to antibiotics and other toxic chemicals. *Trends Microbiol* **7**, 410-3.

Ausubel, F.M., R. Brent, R.E. Kingston, D.D. Moor, J.G. Seidman, J.A. Smith & K. Struhl. 2007. Current protocols in molecular biology. *John Wiley & Sons Inc. NY*.

Becerra, M.C., S.C. Appleton, M.F. Franke, K. Chalco, J. Bayona, M.B. Murray & C.D. Mitnick. 2010. Recurrence after treatment for pulmonary multidrug-resistant tuberculosis. *Clin Infect Dis* **51**, 709-11.

Bentley, S.D., K.F. Chater, A.M. Cerdeno-Tarraga, G.L. Challis, N.R. Thomson, K.D. James, D.E. Harris, M.A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C.W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C.H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O'Neil, E. Rabbinowitsch, M.A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R. Squares, S. Squares, K. Taylor, T. Warren, A. Wietzorrek, J. Woodward, B.G. Barrell, J. Parkhill & D.A. Hopwood. 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**, 141-7.

Berman, H.M., T. Battistuz, T.N. Bhat, W.F. Bluhm, P.E. Bourne, K. Burkhardt, Z. Feng, G.L. Gilliland, L. Iype, S. Jain, P. Fagan, J. Marvin, D. Padilla, V. Ravichandran, B. Schneider, N. Thanki, H. Weissig, J.D. Westbrook & C. Zardecki. 2002. The Protein Data Bank. *Acta Crystallogr D Biol Crystallogr* **58**, 899-907.

Bond, C.S. 2003. TopDraw: a sketchpad for protein structure topology cartoons. *Bioinformatics* **19**, 311-2.

Boucher, H., L.G. Miller & R.R. Razonable. 2010. Serious infections caused by methicillin-resistant *Staphylococcus aureus*. *Clin Infect Dis* **51 Suppl 2**, S183-97.

Buchmeier, N., S. Bossie, C.Y. Chen, F.C. Fang, D.G. Guiney & S.J. Libby. 1997. SlyA, a transcriptional regulator of *Salmonella typhimurium*, is required for

resistance to oxidative stress and is expressed in the intracellular environment of macrophages. *Infect Immun* **65**, 3725-30.

Buntin, K., H. Irschik, K.J. Weissman, E. Luxenburger, H. Blocker & R. Muller. 2010a. Biosynthesis of thuggacins in myxobacteria: comparative cluster analysis reveals basis for natural product structural diversity. *Chem Biol* **17**, 342-56.

Buntin, K., K.J. Weissman & R. Muller. 2010b. An unusual thioesterase promotes isochromanone ring formation in ajudazol biosynthesis. *Chembiochem* **11**, 1137-46.

Cathelyn, J.S., S.D. Crosby, W.W. Lathem, W.E. Goldman & V.L. Miller. 2006. RovA, a global regulator of *Yersinia pestis*, specifically required for bubonic plague. *Proc Natl Acad Sci U S A* **103**, 13514-9.

Cathelyn, J.S., D.W. Ellison, S.J. Hinchliffe, B.W. Wren & V.L. Miller. 2007. The RovA regulons of *Yersinia enterocolitica* and *Yersinia pestis* are distinct: evidence that many RovA-regulated genes were acquired more recently than the core genome. *Mol Microbiol* **66**, 189-205.

Challis, G.L., J. Ravel & C.A. Townsend. 2000. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem Biol* **7**, 211-24.

Chauhan, D., L. Catley, G. Li, K. Podar, T. Hideshima, M. Velankar, C. Mitsiades, N. Mitsiades, H. Yasui, A. Letai, H. Ovaa, C. Berkers, B. Nicholson, T.-H. Chao, S.T.C. Neuteboom, P. Richardson, M.A. Palladino & K.C. Anderson. 2005. A novel orally active proteasome inhibitor induces apoptosis in multiple myeloma cells with mechanisms distinct from Bortezomib. *Cancer Cell* **8**, 407-419.

Chen, V.B., W.B. Arendall, 3rd, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L.W. Murray, J.S. Richardson & D.C. Richardson. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21.

Coligan, J.E., B.M. Dunn, H.L. Ploegh, D.W. Speicher & P.T. Wingfield. 2002. Current protocols in protein science. *John Wiley & Sons Inc., New York*.

Cotelesage, J.J., J. Puttick, H. Goldie, B. Rajabi, B. Novakovski & L.T. Delbaere. 2007. How does an enzyme recognize CO2? *Int J Biochem Cell Biol* **39**, 1204-10.

Cowtan, K. 2006. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* **62**, 1002-11.

Cowtan, K. 2010. Recent developments in classical density modification. *Acta Crystallogr D Biol Crystallogr* **66**, 470-8.

Dame, R.T., M.S. Luijsterburg, E. Krin, P.N. Bertin, R. Wagner & G.J. Wuite. 2005. DNA bridging: a property shared among H-NS-like proteins. *J Bacteriol* **187**, 1845-8.

Davis, I.W., A. Leaver-Fay, V.B. Chen, J.N. Block, G.J. Kapral, X. Wang, L.W. Murray, W.B. Arendall, 3rd, J. Snoeyink, J.S. Richardson & D.C. Richardson. 2007. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**, W375-83.

Deghmane, A.E., D. Giorgini, L. Maigre & M.K. Taha. 2004. Analysis *in vitro* and *in vivo* of the transcriptional regulator CrgA of *Neisseria meningitidis* upon contact with target cells. *Mol Microbiol* **53**, 917-27.

Dube, P. 2009. Interaction of *Yersinia* with the gut: mechanisms of pathogenesis and immune evasion. *Curr Top Microbiol Immunol* **337**, 61-91.

Eble, J.A., K.W. Wucherpfennig, L. Gauthier, P. Dersch, E. Krukonis, R.R. Isberg & M.E. Hemler. 1998. Recombinant soluble human alpha 3 beta 1 integrin: purification, processing, regulation, and specific binding to laminin-5 and invasin in a mutually exclusive manner. *Biochemistry* **37**, 10945-55.

Edwards, K.J., J.D. Barton, J. Rossjohn, J.M. Thorn, G.L. Taylor & D.L. Ollis. 1996. Structural and sequence comparisons of quinone oxidoreductase, zeta-crystallin, and glucose and alcohol dehydrogenases. *Arch Biochem Biophys* **328**, 173-83.

El Tahir, Y. & M. Skurnik. 2001. YadA, the multifaceted *Yersinia* adhesin. *Int J Med Microbiol* **291**, 209-18.

Ellison, D.W., M.B. Lawrenz & V.L. Miller. 2004. Invasin and beyond: regulation of *Yersinia* virulence by RovA. *Trends Microbiol* **12**, 296-300.

Ellison, D.W. & V.L. Miller. 2006. H-NS represses inv transcription in *Yersinia enterocolitica* through competition with RovA and interaction with YmoA. *J Bacteriol* **188**, 5101-12.

Emsley, P., B. Lohkamp, W.G. Scott & K. Cowtan. 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501.

Erb, T.J., I.A. Berg, V. Brecht, M. Muller, G. Fuchs & B.E. Alber. 2007. Synthesis of C5-dicarboxylic acids from C2-units involving crotonyl-CoA carboxylase/reductase: the ethylmalonyl-CoA pathway. *Proc Natl Acad Sci U S A* **104**, 10631-6.

Erb, T.J., V. Brecht, G. Fuchs, M. Muller & B.E. Alber. 2009. Carboxylation mechanism and stereochemistry of crotonyl-CoA carboxylase/reductase, a carboxylating enoyl-thioester reductase. *Proc Natl Acad Sci U S A* **106**, 8871-6.

Esposito, L., I. Bruno, F. Sica, C.A. Raia, A. Giordano, M. Rossi, L. Mazzarella & A. Zagari. 2003. Crystal structure of a ternary complex of the alcohol dehydrogenase from Sulfolobus solfataricus. *Biochemistry* **42**, 14397-407.

Eustaquio, A.S., R.P. McGlinchey, Y. Liu, C. Hazzard, L.L. Beer, G. Florova, M.M. Alhamadsheh, A. Lechner, A.J. Kale, Y. Kobayashi, K.A. Reynolds & B.S. Moore. 2009. Biosynthesis of the salinosporamide A polyketide synthase

substrate chloroethylmalonyl-coenzyme A from S-adenosyl-L-methionine. *Proc Natl Acad Sci U S A* **106**, 12295-300.

Evans, G. & R. Pettifer. 2001. CHOOCH: a program for deriving anomalous-scattering factors from X-ray fluorescence spectra. *J. Appl. Cryst.* **34**, 82-86.

Ezezika, O.C., S. Haddad, T.J. Clark, E.L. Neidle & C. Momany. 2007. Distinct effector-binding sites enable synergistic transcriptional activation by BenM, a LysR-type regulator. *J Mol Biol* **367**, 616-29.

Frank, B., J. Knauber, H. Steinmetz, M. Scharfe, H. Blocker, S. Beyer & R. Muller. 2007. Spiroketal polyketide formation in Sorangium: identification and analysis of the biosynthetic gene cluster for the highly cytotoxic spirangienes. *Chem Biol* **14**, 221-33.

Fuangthong, M., S. Atichartpongkul, S. Mongkolsuk & J.D. Helmann. 2001. OhrR is a repressor of ohrA, a key organic hydroperoxide resistance determinant in *Bacillus subtilis*. *J Bacteriol* **183**, 4134-41.

Fuangthong, M. & J.D. Helmann. 2002. The OhrR repressor senses organic hydroperoxides by reversible formation of a cysteine-sulfenic acid derivative. *Proc Natl Acad Sci U S A* **99**, 6690-5.

Galimand, M., A. Guiyoule, G. Gerbaud, B. Rasoamanana, S. Chanteau, E. Carniel & P. Courvalin. 1997. Multidrug Resistance in *Yersinia pestis* Mediated by a Transferable Plasmid. *New England Journal of Medicine* **337**, 677-681.

Garman, E.F. 2010. Radiation damage in macromolecular crystallography: what is it and why should we care? *Acta Crystallogr D Biol Crystallogr* **66**, 339-51.

Gasteiger, E., A. Gattiker, C. Hoogland, I. Ivanyi, R.D. Appel & A. Bairoch. 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* **31**, 3784-8.

Gokhale, R.S., R. Sankaranarayanan & D. Mohanty. 2007. Versatility of polyketide synthases in generating metabolic diversity. *Curr Opin Struct Biol* **17**, 736-43.

Gouet, P., E. Courcelle, D.I. Stuart & F. Metoz. 1999. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305-8.

Haider, F., J.K. Lithgow, M.R. Stapleton, V.A. Norte, R.E. Roberts & J. Green. 2008. DNA recognition by the *Salmonella enterica serovar Typhimurium* transcription factor SlyA. *Int Microbiol* **11**, 245-50.

Harris, S. 1994. Factories of death. *New York, NY: Routledge* **74**, 96.

Heesemann, J., A. Sing & K. Trulzsch. 2006. *Yersinia's* stratagem: targeting innate and adaptive immune defense. *Curr Opin Microbiol* **9**, 55-61.

Herbst, K., M. Bujara, A.K. Heroven, W. Opitz, M. Weichert, A. Zimmermann & P. Dersch. 2009. Intrinsic thermal sensing controls proteolysis of *Yersinia* virulence regulator RovA. *PLoS Pathog* **5**, e1000435.

Heroven, A.K., K. Bohme, M. Rohde & P. Dersch. 2008. A Csr-type regulatory system, including small non-coding RNAs, regulates the global virulence regulator RovA of *Yersinia pseudotuberculosis* through RovM. *Mol Microbiol* **68**, 1179-95.

Heroven, A.K., K. Bohme, H. Tran-Winkler & P. Dersch. 2007. Regulatory elements implicated in the environmental control of invasin expression in enteropathogenic *Yersinia*. *Adv Exp Med Biol* **603**, 156-66.

Heroven, A.K. & P. Dersch. 2006. RovM, a novel LysR-type regulator of the virulence activator gene *rovA*, controls cell invasion, virulence and motility of *Yersinia pseudotuberculosis*. *Mol Microbiol* **62**, 1469-83.

Heroven, A.K. & P. Dersch. 2010. A global network, implicating regulatory RNAs, controls virulence gene expression in pathogenic *yersiniae*. *Res. Adv. in Molecular Microbiology* **1**, 1-17.

Heroven, A.K., G. Nagel, H.J. Tran, S. Parr & P. Dersch. 2004. RovA is autoregulated and antagonizes H-NS-mediated silencing of invasin and rovA expression in *Yersinia pseudotuberculosis*. *Mol Microbiol* **53**, 871-88.

Hirsch, E.B. & V.H. Tam. 2010. Impact of multidrug-resistant *Pseudomonas aeruginosa* infection on patient outcomes. *Expert Rev Pharmacoecon Outcomes Res* **10**, 441-51.

Hong, M., M. Fuangthong, J.D. Helmann & R.G. Brennan. 2005. Structure of an OhrR-ohrA operator complex reveals the DNA binding mechanism of the MarR family. *Mol Cell* **20**, 131-41.

Inglesby, T.V., D.T. Dennis, D.A. Henderson, J.G. Bartlett, M.S. Ascher, E. Eitzen, A.D. Fine, A.M. Friedlander, J. Hauer, J.F. Koerner, M. Layton, J. McDade, M.T. Osterholm, T. O'Toole, G. Parker, T.M. Perl, P.K. Russell, M. Schoch-Spana & K. Tonat. 2000. Plague as a biological weapon: medical and public health management. Working Group on Civilian Biodefense. *Jama* **283**, 2281-90.

Jiang, J., X. He & D.E. Cane. 2006. Geosmin biosynthesis. *Streptomyces coelicolor* germacradienol/germacrene D synthase converts farnesyl diphosphate to geosmin. *J Am Chem Soc* **128**, 8128-9.

Kabsch, W. 2010. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-32.

Khosla, C., S. Kapur & D.E. Cane. 2009. Revisiting the modularity of modular polyketide synthases. *Curr Opin Chem Biol* **13**, 135-43.

Knapp, G.S. & J.C. Hu. 2009. The oligomerization of CynR in Escherichia coli. *Protein Sci* **18**, 2307-15.

Koornhof, H.J., R.A. Smego, Jr. & M. Nicol. 1999. Yersiniosis. II: The pathogenesis of *Yersinia* infections. *Eur J Clin Microbiol Infect Dis* **18**, 87-112.

Krissinel, E. & K. Henrick. 2007. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-97.

Kumarevel, T., T. Tanaka, T. Umehara & S. Yokoyama. 2009. ST1710-DNA complex crystal structure reveals the DNA binding mechanism of the MarR family of regulators. *Nucleic Acids Res* **37**, 4723-35.

Lai, J.R., A. Koglin & C.T. Walsh. 2006. Carrier protein structure and recognition in polyketide and nonribosomal peptide biosynthesis. *Biochemistry* **45**, 14869-79.

Lambalot, R.H., A.M. Gehring, R.S. Flugel, P. Zuber, M. LaCelle, M.A. Marahiel, R. Reid, C. Khosla & C.T. Walsh. 1996. A new enzyme superfamily - the phosphopantetheinyl transferases. *Chem Biol* **3**, 923-36.

Larkin, M.A., G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson & D.G. Higgins. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-8.

Lerat, S., A.-M. Simao-Beaunoir & C. Beaulieu. 2009. Genetic and physiological determinants of *Streptomyces scabies* pathogenicity. *Molecular Plant Pathology* **10**, 579-585.

Li, X., S. Jayachandran, H.H. Nguyen & M.K. Chan. 2007. Structure of the *Nitrosomonas europaea* Rh protein. *Proc Natl Acad Sci U S A* **104**, 19279-84.

Lim, D., K. Poole & N.C. Strynadka. 2002. Crystal structure of the MexR repressor of the mexRAB-oprM multidrug efflux operon of *Pseudomonas aeruginosa*. *J Biol Chem* **277**, 29253-9.

Maier, T., M. Leibundgut & N. Ban. 2008. The crystal structure of a mammalian fatty acid synthase. *Science* **321**, 1315-22.

Martin, R.G. & J.L. Rosner. 1995. Binding of purified multiple antibiotic-resistance repressor protein (MarR) to mar operator sequences. *Proc Natl Acad Sci U S A* **92**, 5456-60.

Matsumoto, H. & G.M. Young. 2009. Translocated effectors of *Yersinia*. *Curr Opin Microbiol* **12**, 94-100.

Matthews, B.W. 1968. Some crystal forms of bovine chymotrypsinogen B and chymotrypsinogen A. *J Mol Biol* **33**, 499-501.

McCoy, A.J., R.W. Grosse-Kunstleve, P.D. Adams, M.D. Winn, L.C. Storoni & R.J. Read. 2007. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674.

Monferrer, D., T. Tralau, M.A. Kertesz, I. Dix, M. Sola & I. Uson. 2010. Structural studies on the full-length LysR-type regulator TsaR from *Comamonas testosteroni* T-2 reveal a novel open conformation of the tetrameric LTTR fold. *Mol Microbiol* **75**, 1199-214.

Muraoka, S., R. Okumura, N. Ogawa, T. Nonaka, K. Miyashita & T. Senda. 2003. Crystal structure of a full-length LysR-type transcriptional regulator, CbnR: unusual combination of two subunit forms and molecular bases for causing and changing DNA bend. *J Mol Biol* **328**, 555-66.

Murshudov, G.N., A.A. Vagin & E.J. Dodson. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240-55.

Nagel, G., A. Lahrz & P. Dersch. 2001. Environmental control of invasin expression in *Yersinia pseudotuberculosis* is mediated by regulation of RovA, a transcriptional activator of the SlyA/Hor family. *Mol Microbiol* **41**, 1249-69.

Niggemann, J., N. Bedorf, U. Flörke, H. Steinmetz, K. Gerth, H. Reichenbach & G. Höfle. 2005. Spirangien A and B, Highly Cytotoxic and Antifungal Spiroketals from the Myxobacterium *Sorangium cellulosum*: Isolation, Structure Elucidation and Chemical Modifications. *European Journal of Organic Chemistry* **2005**, 5013-5018.

Noom, M.C., W.W. Navarre, T. Oshima, G.J. Wuite & R.T. Dame. 2007. H-NS promotes looped domain formation in the bacterial chromosome. *Curr Biol* **17**, R913-4.

Notredame, C. 2010. Computing multiple sequence/structure alignments with the T-coffee package. *Curr Protoc Bioinformatics* **Chapter 3**, Unit 3 8 1-25.

Ollis, D.L., E. Cheah, M. Cygler, B. Dijkstra, F. Frolow, S.M. Franken, M. Harel, S.J. Remington, I. Silman, J. Schrag & et al. 1992. The alpha/beta hydrolase fold. *Protein Eng* **5**, 197-211.

Painter, J. & E.A. Merritt. 2006a. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D Biol Crystallogr* **62**, 439-50.

Painter, J. & E.A. Merritt. 2006b. TLSMD webserver for the generation of multi-groupTLS models. *Journal of Applied Crystallography* **39**, 109-111.

Pandey, A.S., D.W. Mulder, S.A. Ensign & J.W. Peters. 2011. Structural basis for carbon dioxide binding by 2-ketopropyl coenzyme M oxidoreductase/carboxylase. *FEBS Lett* **585**, 459-64.

Perry, R.D. & J.D. Fetherston. 1997. *Yersinia pestis*--etiologic agent of plague. *Clin Microbiol Rev* **10**, 35-66.

Porte, S., E. Valencia, E.A. Yakovtseva, E. Borras, N. Shafqat, J.E. Debreczeny, A.C. Pike, U. Oppermann, J. Farres, I. Fita & X. Pares. 2009. Three-dimensional structure and enzymatic function of proapoptotic human p53-inducible quinone oxidoreductase PIG3. *J Biol Chem* **284**, 17194-205.

Providenti, M.A. & R.C. Wyndham. 2001. Identification and functional characterization of CbaR, a MarR-like modulator of the cbaABC-encoded chlorobenzoate catabolism pathway. *Appl Environ Microbiol* **67**, 3530-41.

Quade, N., M. Dieckmann, M. Haffke, A.K. Heroven, P. Dersch & D.W. Heinz. 2011. Structure of the effector-binding domain of the LysR-type transcription factor RovM from *Yersinia pseudotuberculosis*. *Acta Crystallographica Section D* **67**, 81-90.

Quintana, E., K. Wierzbicka, P. Mackiewicz, A. Osman, A. Fahal, M. Hamid, J. Zakrzewska-Czerwinska, L. Maldonado & M. Goodfellow. 2008. *Streptomyces sudanensis sp. nov.,* a new pathogen isolated from patients with actinomycetoma. *Antonie van Leeuwenhoek* **93**, 305-313.

Rachid, S., L. Huo, J. Herrmann, M. Stadler, B. Kopcke, J. Bitzer & R. Muller. 2011. Mining the cinnabaramide biosynthetic pathway to generate novel proteasome inhibitors. *Chembiochem* **12**, 922-31.

Ramachandran, G.N. & V. Sasisekharan. 1968. Conformation of polypeptides and proteins. *Adv Protein Chem* **23**, 283-438.

Revell, P.A. & V.L. Miller. 2000. A chromosomally encoded regulator is required for expression of the *Yersinia enterocolitica inv* gene and for virulence. *Mol Microbiol* **35**, 677-85.

Ruangprasert, A., S.H. Craven, E.L. Neidle & C. Momany. 2010. Full-Length Structures of BenM and Two Variants Reveal Different Oligomerization Schemes for LysR-Type Transcriptional Regulators. *J Mol Biol*.

Sainsbury, S., L.A. Lane, J. Ren, R.J. Gilbert, N.J. Saunders, C.V. Robinson, D.I. Stuart & R.J. Owens. 2009. The structure of CrgA from *Neisseria meningitidis* reveals a new octameric assembly state for LysR transcriptional regulators. *Nucleic Acids Res* **37**, 4545-58.

Samel, S.A., B. Wagner, M.A. Marahiel & L.O. Essen. 2006. The thioesterase domain of the fengycin biosynthesis cluster: a structural base for the macrocyclization of a non-ribosomal lipopeptide. *J Mol Biol* **359**, 876-89.

Saridakis, V., D. Shahinas, X. Xu & D. Christendat. 2008. Structural insight on the mechanism of regulation of the MarR family of proteins: high-resolution crystal structure of a transcriptional repressor from *Methanobacterium thermoautotrophicum*. *J Mol Biol* **377**, 655-67.

Scaglione, J.B., D.L. Akey, R. Sullivan, J.D. Kittendorf, C.M. Rath, E.S. Kim, J.L. Smith & D.H. Sherman. 2010. Biochemical and structural characterization of the tautomycetin thioesterase: analysis of a stereoselective polyketide hydrolase. *Angew Chem Int Ed Engl* **49**, 5726-30.

Schell, M.A. 1993. Molecular biology of the LysR family of transcriptional regulators. *Annu Rev Microbiol* **47**, 597-626.

Schuttelkopf, A.W. & D.M. van Aalten. 2004. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr* **60**, 1355-63.

Sheldrick, G.M. 2008. A short history of SHELX. *Acta Crystallogr A* **64**, 112-22.

Sjoblom, B., M. Polentarutti & K. Djinovic-Carugo. 2009. Structural study of X-ray induced activation of carbonic anhydrase. *Proc Natl Acad Sci U S A* **106**, 10609-13.

Smirnova, I.A., C. Dian, G.A. Leonard, S. McSweeney, D. Birse & P. Brzezinski. 2004. Development of a bacterial biosensor for nitrotoluenes: the crystal structure of the transcriptional regulator DntR. *J Mol Biol* **340**, 405-18.

Spivey, H.O. & J. Ovadi. 1999. Substrate channeling. *Methods* **19**, 306-21.

Stadler, M., J. Bitzer, A. Mayer-Bartschmid, H. Müller, J. Benet-Buchholz, F. Gantner, H.-V. Tichy, P. Reinemer & K.B. Bacon. 2007. Cinnabaramides A-G:Analogues of Lactacystin and Salinosporamide from a Terrestrial *Streptomycete*. *Journal of Natural Products* **70**, 246-252.

Stapleton, M.R., V.A. Norte, R.C. Read & J. Green. 2002. Interaction of the *Salmonella typhimurium* transcription and virulence factor SlyA with target DNA and identification of members of the SlyA regulon. *J Biol Chem* **277**, 17630-7.

Stenseth, N.C., B.B. Atshabar, M. Begon, S.R. Belmain, E. Bertherat, E. Carniel, K.L. Gage, H. Leirs & L. Rahalison. 2008. Plague: past, present, and future. *PLoS Med* **5**, e3.

Sulakvelidze, A. 2000. *Yersiniae* other than *Y. enterocolitica*, *Y. pseudotuberculosis*, and *Y. pestis*: the ignored species. *Microbes Infect* **2**, 497-513.

Tan, S., Y. Hunziker, L. Pellegrini & T.J. Richmond. 2000. Crystallization of the yeast MAT[alpha]2/MCM1/DNA ternary complex: general methods and principles for protein/DNA cocrystallization. *Journal of Molecular Biology* **297**, 947-959.

Tauxe, R.V. 2004. Salad and pseudoappendicitis: *Yersinia pseudotuberculosis* as a foodborne pathogen. *J Infect Dis* **189**, 761-3.

Thompson, C.J., D. Fink & L.D. Nguyen. 2002. Principles of microbial alchemy: insights from the *Streptomyces coelicolor* genome sequence. *Genome Biol* **3**, REVIEWS1020.

Tran, H.J., A.K. Heroven, L. Winkler, T. Spreter, B. Beatrix & P. Dersch. 2005. Analysis of RovA, a transcriptional regulator of *Yersinia pseudotuberculosis* virulence that acts through antirepression and direct transcriptional activation. *J Biol Chem* **280**, 42423-32.

Trosky, J.E., A.D. Liverman & K. Orth. 2008. *Yersinia* outer proteins: Yops. *Cell Microbiol* **10**, 557-65.

Trulzsch, K., M.F. Oellerich & J. Heesemann. 2007. Invasion and dissemination of *Yersinia enterocolitica* in the mouse infection model. *Adv Exp Med Biol* **603**, 279-85.

Tsai, S.C., L.J. Miercke, J. Krucinski, R. Gokhale, J.C. Chen, P.G. Foster, D.E. Cane, C. Khosla & R.M. Stroud. 2001. Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: versatility from a unique substrate channel. *Proc Natl Acad Sci U S A* **98**, 14808-13.

Verdonk, M.L., G. Chessari, J.C. Cole, M.J. Hartshorn, C.W. Murray, J.W. Nissink, R.D. Taylor & R. Taylor. 2005. Modeling water molecules in protein-ligand docking using GOLD. *J Med Chem* **48**, 6504-15.

Weissman, K.J. & R. Muller. 2008. Protein-protein interactions in multienzyme megasynthetases. *Chembiochem* **9**, 826-48.

Wilkinson, B. & J. Micklefield. 2007. Mining and engineering natural-product biosynthetic pathways. *Nat Chem Biol* **3**, 379-86.

Wilkinson, S.P. & A. Grove. 2004. HucR, a novel uric acid-responsive member of the MarR family of transcriptional regulators from *Deinococcus radiodurans*. *J Biol Chem* **279**, 51442-50.

Wilkinson, S.P. & A. Grove. 2006. Ligand-responsive transcriptional regulation by members of the MarR family of winged helix proteins. *Curr Issues Mol Biol* **8**, 51-62.

Wren, B.W. 2003. The *yersiniae*--a model genus to study the rapid evolution of bacterial pathogens. *Nat Rev Microbiol* **1**, 55-64.

Wu, R.Y., R.G. Zhang, O. Zagnitko, I. Dementieva, N. Maltzev, J.D. Watson, R. Laskowski, P. Gornicki & A. Joachimiak. 2003. Crystal structure of Enterococcus faecalis SlyA-like transcriptional factor. *J Biol Chem* **278**, 20240-4.

Yang, F., Y. Ke, Y. Tan, Y. Bi, Q. Shi, H. Yang, J. Qiu, X. Wang, Z. Guo, H. Ling, R. Yang & Z. Du. 2010. Cell membrane is impaired, accompanied by enhanced type III secretion system expression in *Yersinia pestis* deficient in RovA regulator. *PLoS One* **5**.

Zhou, D. & R. Yang. 2009. Molecular Darwinian evolution of virulence in *Yersinia pestis*. *Infect Immun* **77**, 2242-50.

Zhou, X., Z. Lou, S. Fu, A. Yang, H. Shen, Z. Li, Y. Feng, M. Bartlam, H. Wang & Z. Rao. 2010. Crystal structure of ArgP from *Mycobacterium tuberculosis* confirms two distinct conformations of full-length LysR transcriptional regulators and reveals its function in DNA binding and transcriptional regulation. *J Mol Biol* **396**, 1012-24.

Danksagung

Bei Prof. Dr. Dirk Heinz bedanke ich mich herzlich, dass ich in seiner Gruppe unter hervorragenden Bedingungen an spannenden Themen arbeiten durfte. In unseren Diskussionen half er mir mit seiner Erfahrung und seinem Wissen weiter und hatte so entscheidenden Anteil an dem Erfolg meiner Arbeit. Weiterhin bin ich ihm wegen seiner Unterstützung, was meinen weiteren Werdegang anbelangt, und wegen der Möglichkeit eine Vielzahl von internationalen Konferenzen zu besuchen, um Kontakte zu knüpfen und meinen Horizont zu erweitern, zu großem Dank verpflichtet.

Prof. Dr. Michael Steinert danke ich sehr herzlich für die bereitwillige Übernahme des Zweitgutachtens.

Bei Prof. Dr. Dieter Jahn bedanke ich mich herzlich für die Übernahme des Prüfungsvorsitzes.

Prof. Dr. Petra Dersch danke ich für die fruchtbare Kooperation und die vielen hilfreichen Diskussionen. Wann immer ich Fragen hatte oder Hilfe brauchte, war sie sofort zur Stelle.

Bei der ganzen Arbeitsgruppe MIBI und insbesondere bei Dr. Ann Kathrin Heroven, Dr. Katharina Herbst und Chriselle Mendonca möchte ich mich bedanken für die Kooperation, die vielen netten und konstruktiven Diskussionen und dafür, dass die Gruppe für mich eine Art zweite Heimat geworden ist.

Prof. Dr. Rolf Müller und seinen Mitarbeitern Dr. Rashid Shwan und Liujie Huo danke ich für die gute Kooperation und die vielen Hilfestellung bei unseren Projekten.

Ganz besonders möchte ich mich zudem bei Dr. Joachim Reichelt, Dr. Jörn Kraußе und Dr. Björn Klink bedanken, die mir immer wieder sehr geholfen haben bei der Auswertung meiner Daten und den zahlreichen kleinen Computerproblemen.

Außerdem möchte mich ganz herzlich bei meinen Diplomanden Marlene Frank und Christoph Bürth sowie bei meiner Bacherlorandin Marieke Dieckmann und meinem Praktikanten Johannes-Urban Mayer für die großartige tatkräftige Unterstützung bedanken. Ohne sie hätte ich sicherlich nur die Hälfte von dem erreicht, was sich hier in meiner Arbeit findet.

Bei Prof. Dr. Christiane Ritter möchte ich mich herzlich für die Hilfe bei NMR bedanken.

Ein ganz großes Dankeschön geht an das ganze MOSB-Team für die tolle Arbeitsathmosphäre, die vielen, vielen kleinen Hilfestellungen und natürlich auch die Unterhaltungen, sowohl wissenschaftlicher als sonstiger Natur.

Weiterhin will ich mich bei meinen Eltern und meiner Freundin Julia Bremer bedanken, die mir mit ihrer Liebe und ihrer moralischen und manchmal auch tatkräftigen Unterstützung durch manche schwierige Zeit geholfen haben.

Lebenslauf

Persönliche Angaben

Nick Quade

Geb. am 06.06.1985 in Hannover

Ledig

Deutsch

Schulischer Werdegang

1991-1992 Grundschule Goetheschule, Hannover

1992-1995 Grundschule Friesenstraße, Hannover

1995-1997 Orientierungsstufe Lührstraße, Hannover

1997-2003 Kaiser Wilhelm- und Ratsgymnasium, Hannover (die 10. Klasse wurde übersprungen)

2003 Abitur am Kaiser Wilhelm- und Ratsgymnasium, Hannover

Akademischer Werdegang

2003-2008 Studium der Biochemie auf Diplom an der Leibniz Universität Hannover

2007-2008 Diplomarbeit in der Arbeitsgruppe von Dr. Xiaodong Zhang am Imperial College London

Seit 2008 Promotion am Helmholtz Zentrum für Infektionsforschung, Braunschweig, im Bereich Strukturbiologie in der Arbeitsgruppe von Prof. Dr. Dirk Heinz.